INVESTIGATING THE REPAIR AND SITE SELECTIVE FORMATION OF INTERSTRAND CROSS-LINKS USING SYNTHETIC ANALOGUES DERIVED FROM OXIDIZED ABASIC SITES

by

Souradyuti Ghosh

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

24th March, 2015

Abstract

Interstrand cross-links (ICLs) are covalent linkages formed by reaction of nucleobases with a variety of bifunctional agents of exogenous (e.g. nitrogen mustard reagents) and endogenous (e.g. malondialdehyde) origin. Recently, electrophilic DNA lesions such as an AP site (apurinic/apyrimidinic site), C4'-oxidized abasic site (C4-AP), and dioxobutane (DOB) have also been implicated in ICL formation. A covalent linkage between the DNA strands prevents their separation, acting as a block for necessary cellular processes such as transcription and replication. Given the deleterious consequences of ICL formation on cellular pathways, eukaryotes and prokaryotes use nucleotide excision repair (NER) to remove interstrand cross-links. On one hand, cross-link-causing external agents, a number of which are active anti-cancer drugs, try to exploit the toxicity of ICLs to kill the cell. On the other hand, diseases that hamper activities of proteins involved in NER (e.g. Fanconi anemia) increase vulnerability to ICL-causing agents. Therefore, understanding the mechanism of cross-link formation and tracing their biochemical fate remains extremely important.

C4-AP and DOB lesions are formed by anti-cancer chemotherapeutics and γ-irradiation. The lesions reversibly react with the N6 amine of an adenosine (A) to form an unstable ICL. While the ICL between C4-AP and A has 3 h half-life, the one between DOB and A is modestly more stable ($t_{1/2}$ = 10.1 h). Although unstable, both ICLs last long enough to potentially disrupt replication and transcription in cells. Intriguingly, both C4-AP and DOB ICLs are formed in a site selective manner, with an A opposite their 3'-adjacent thymidine (T). The distances computed between N6 amine of A at the preferred site and the C1'-aldehyde of the electrophilic lesions is not radically different from that

involving an opposing A, making the site selectivity ambiguous. The C4-AP ICL eventually reverts to the lesion, undergoes a strand scission and forms a thermodynamic cross-link. The short lifetime of DOB and C4-AP ICLs make the investigation of their biochemical fate and site selective formation difficult.

Herein, we report the development of synthetic methodologies to prepare stable analogues of DOB and C4-AP ICLs. The procedures involve orthogonal removal of protecting groups from oligonucleotides on solid support without affecting any native DNA nucleobase or cyanoethyl protection and subsequent solid phase synthesis. In addition to being able to prepare any cross-link irrespective of sequence, the strategies can be applied to the synthesis of branched DNA and RNA.

The stable analogues of C4-AP and DOB ICLs were treated with UvrABC, the bacterial nucleotide excision repair proteins to understand their repair. The incision on C4-AP lesion was slow, with maximum 27 – 30% cleavage reported after 8 h, implying potentially deleterious consequence for processes like transcription and replication. No double strand break caused by repetitive incision on opposing strand was detected. In contrast, the DOB ICL analogue exclusively produced double strand breaks upon UvrABC treatment.

In addition to distance, relative stability in resulting ICL secondary structure was hypothesized as a driving force behind site selective cross-link formation. Stable oligonucleotides analogues of DOB and C4-AP ICLs with “observed” (ICL with A opposite 3'-adjacent T) and “unobserved” (ICL with opposing A) connectivities were synthesized. Upon treatment with chemical agents that detect distorted (solvent exposed)

nucleotides, the oligonucleotides with "unobserved" ICLs were found to have greater distortion. This was further verified using thermodynamic UV melting, where the "observed" ICLs had higher T_ms.

In conclusion, our work describes novel strategies for synthesizing stable analogues of DOB and C4-AP ICLs irrespective of their sequence. The work is compatible with other ICL preparations, but is also extendable to branched DNA and RNA synthesis. The stable analogues were employed to address the biochemical fate of the respective ICLs in the context of bacterial nucleotide excision repair proteins, UvrABC endonuclease. Finally, the structures of the analogues were probed using chemical agents. The result was correlated with thermodynamic UV melting to draw light on the intriguing site selective formation of native C4-AP and DOB ICLs.

Advisor:	Prof. Marc M. Greenberg
Readers:	Prof. John P. Toscano
	Prof. Craig A. Townsend

Acknowledgements

This work would not have been completed without the help of a number of people. Thank you my parents, Subhash Chandra Ghosh and Namita Ghosh for their support throughout these five and a half years. There has been times when they missed me and needed me by their side in India. But having understood my requirements with the degree program in JHU, they have never demanded anything from me, but always extended their support when I needed it most. Thanks to my brother, Saraswata Ghosh, who himself being an engineer, has been an constant inspiration and source of latitudes of information for me. Thanks to Anuradha, my wife, who got married to me in June, 2013. Despite the time 2013 – 2015 being the busiest time of my grad life in JHU, she was never impatient, angry and has never given up on me. She has been an active support in my research life, be it accompanying me to lab at odd hours or providing constant inspiration during the thesis-writing period. It could not have been completed without her.

Thank you Marc (Prof. Greenberg), my advisor for the last 5.5 years, for your constant efforts to develop me into a better scientist, and an able, mature and (probably) a more organized human being. This spans every aspects of my academic life in JHU, ranging from coursework to experiments writing to presentation to literature search to managing and fixing lab equipments and what not. Your tireless efforts to make me a better experimental scientist has immensely benefitted me here, and will serve as a beacon in my future endeavors.

Thanks to the past and present members of Greenberg lab, Marino and Dianje (for teaching synthetic organic experiment and for sharing adventures and misadventures),

Marisa, Chuanzheng, Lewei, Joanna and John (for introducing me to the world of biochemistry and for the memorable friendship), Jack Sloane (for sharing bench space, bad jokes and the “cool” NCB 336), Jonathan T. Sczepanski, Lirui Guan and Aaron Jacobs (for their helpful discussion), Rakesh, Arnab, (for the discussions, insights and encouragements) Dumitru, Supratim, Josh, (Lord) Lewei, and Danny (for their awesome support). Thank you Adam, Kristine, Surekha, Preeti, and Daniel. I will always miss how close we all have been in this lab.

Thank you Prof. Toscano and Prof. Townsend, for being my thesis readers and being in my orals and GBO, along with Prof. Miller and Prof. Stivers. Thank you Prof. Draper, Prof. Meyer, and Prof. Tovar, for sharing their knowledge during coursework. Thank you Dr. Phil Mortimer, for the assistance during oligonucleotide MS and MALDI-TOF experiments and for being a constant source of knowledge. Thank you Dr. Cathy Moore and Dr. Chuck Long for the help in NMR experiments. Thank you Jean and Rosalie, for helping with the administrative work, and also for the support and encouragements. Thank you Boris, for helping me fixing equipments, ordering supplies, and be the great friendship. Thank you my classmates in JHU, Pearl, Imtiaz, Chad, Xingxing, Benjamin, Brian, and Nicole.

Thank you to Ayan, Subhabrata, Biola, Moiz. Apart from being fun roommates, friends and mentors. You guys have been and will remain great friends and source of support. Thanks to my highly helpful and knowledgeable friends in Baltimore, especially Rahul, Snehashis, Amarnath, Subrata, Aroop, Krishnarjun, Jaya, Arnab (Mukherjee and Gupta), and Priyasmita for providing essential information throughout my Ph.D life. Thanks to my great friends in Baltimore Sandip, Sumit, Parichoy, Suman, Debmalya,

Sayantan, Taraswi, Sharmistha, Devdatta, Asima. Thanks to my family friends Saurabh and Jayeeta, Jhumjhumi, Rita-kaki and Tuku-kaku. Thanks to my in-laws in India.

Table of Contents

Appendix

List of Tables

List of Figures

List of Schemes

List of Abbreviations

ICL	interstrand cross-link
DNA	deoxyribonucleic acid
AP	abasic site
C4-AP	C4'-oxidized abasic site
DOB	dioxobutane
A	2'-deoxyadenosine
G	2'-deoxyguanosine
C	2'-deoxycytosine
T	2'-deoxythymidine
U	uridine
NER	nucleotide excision repair
DSB	double strand break
NM	nitrogen mustard
CENU	cyanoethylnitrosourea
EMA	N,N-bis(2-chloroethyl)methylamine
DTT	dithiothreitol
MDA	malondialdehyde
γ-OH-PdG	γ-hydroxy-1,N2-propanodeoxyguanosine
ESI-MS	electrospray ionization mass spectrometry
TEA	triethylamine
NV	ortho-nitrobenzyloxycarbonyl
IPP	2,3-dihydro-pyrrolizine bis(isopropylcarbamate)
AVP	2-amino-6-vinylpurine
PAGE	polyacrylamide gel electrophoresis
MC	mitomycin C
T_m	thermodynamic melting
NOESY	nuclear overhauser effect spectroscopy
HMT	4'-(hydroxymethyl)-4.5',8-trimethylpsoralen
BA	benzylhydroxylamine-modified apyrimidine
Lev	levulinyl
ALLOC	allyloxycarbonyl
CE	cyanoethyl

HP	hydroxypyrrolidine
THF	tetradydrofuran
DMAP	dimethylaminopyridine
DMF	dimethylformamide
EDCI	1-ethyl-3-(3-dimethylaminopropyl)carbodiimide
Ac	acetyl
DMT	4,4'-dimethoxytrityl
TLC	thin layer chromatography
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time of flight mass spectrometry
Bz	benzoyl
iBu	isobutyryl
DCM	dichloromethane
DIPEA	diisopropylethylamine
NC	normalized cleavage
DEPC	Diethylpyrocarbonate
DMS	dimethylsulfate
UV	ultraviolet
PIPES	piperazine-N,N′-bis(2-ethanesulfonic acid)
HPLC	high performance liquid chromatography
dNTP	deoxynucleotide triphosphate
EDTA	ethylene diamine tetraacetic acid
LC-MS	liquid chromatography-mass spectrometry
MS	mass spectrometry
NMR	nuclear magnetic resonance
Tris	tris(hydroxymethyl)aminomethane
•OH	hydroxyl radical
NEB	New England Biolab

1. Introduction

DNA interstrand cross-link (ICL) is the formation of a covalent linkage between the strands. The covalent linkages can occur because of reaction between bifunctional agents and nucleobases from opposing strands (Figure 1A). The bifunctional agents have both exogenous (nitrogen mustard reagents,[1] nitrosoureas,[2] mitomycin C,[3] psoralen[4]) and endogenous (malondialdehyde[5]) origin (Figure 1B). Alternatively, reactions between electrophilic lesions such as an AP site,[6,7] C4′-oxidized abasic site (C4-AP),[8] dioxobutane

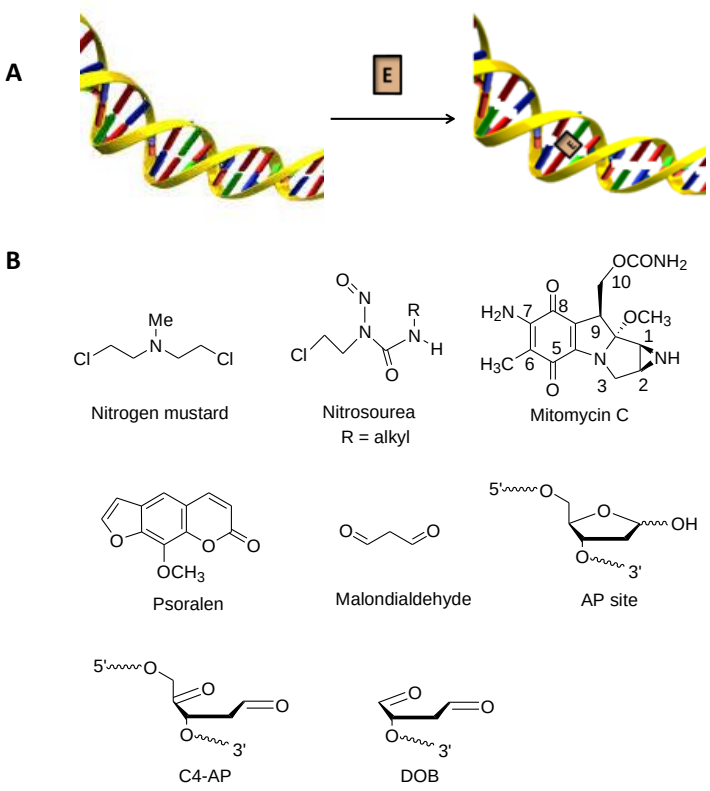


Figure 1. Formation of interstrand cross-link (ICL) by reaction of DNA with a reagent (E) and structure of the AP site, C4-AP and DOB lesions. (A) ICL formation by reaction of DNA with a reagent (E). (B) Chemical structure of the AP site, C4-AP, DOB, nitrogen mustard, nitrosoureas, mitomycin C, psoralen, and malondialdehyde.

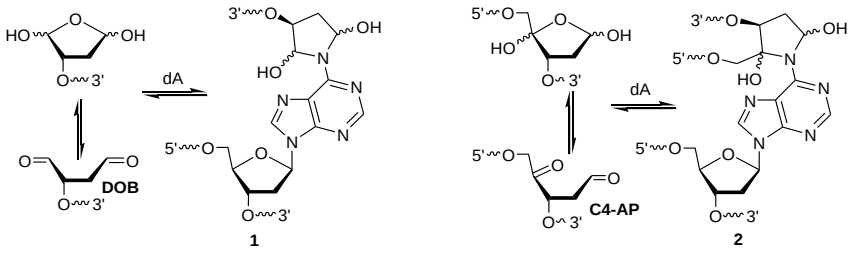
(DOB)[9] and nucleobases can also contribute to ICL formation (Figure 1B). Unless repaired, ICLs hinder cellular processes involving DNA strand separation such as replication[10] and transcription.[11] ICLs are also reported to cause chromosomal breakage,[12]

mutation,[13] and recombigenic effects.[14] Exploiting their damaging potential, nitrogen mustards, mitomycin C and psoralen are used as chemotherapeutic drugs against leukemia,[15] gastrointestinal tumors,[16] and skin diseases[17] respectively. However, low selectivity of these chemotherapeutic agents against cancer cells along with metabolic biogenesis of bis-electrophiles such as malondialdehyde can lead to ICL formation in non-cancerous cells as well, endangering their survival. For these reasons, investigation of the formation and biochemical fate of ICLs is a vibrant area of research.

Due to their potential genotoxicity, ICLs are removed by the nucleotide excision repair (NER) pathway in bacterial and mammalian cells. In bacterial cells UvrA, UvrB, and UvrC endonuclease proteins are responsible for the removal of cross-links. UvrA and UvrB recognize the damage while UvrC carries out the incision.[18] The incision removes a 12 – 13 nucleotide long segment that contains one of the cross-linked nucleotides. Following "gap filling" by a DNA polymerase in that strand, the remnants of the ICL are then removed by incision of the other strand.[18] The incision efficiency is influenced by the size of the cross-link and the deformities it induces in the surrounding nucleotides.[19] However, UvrABC does not always remove ICLs in a flawless manner. Generation of double strand breaks (DSBs) by UvrABC treatment has been reported by simultaneous incision on both strands in more than one instance.[20,21] In another study, DSBs are generated as a consequence of indiscriminate incisions on an ICL accompanied by a strand break.[22] In-order-to capture the complete picture of ICL incision by UvrABC, the structural characterization of the ICL and an understanding of the structural influence on incision are necessary.

The C4′-oxidized abasic site (C4-AP) and dioxobutane (DOB) lesions are derived from hydrogen atom abstraction at the C4′ and C5′-position, respectively. Such hydrogen atom abstractions occur as a result of γ–radiolysis.[23,24] In addition, the antitumor antibiotic bleomycin produces C4-AP, while the enediyne class of chemotherapeutic drugs produces C4-AP and DOB lesions.[25] Both lesions form ICLs with deoxyadenosine (A) (Scheme 1).[8,9] The cross-link by C4-AP is a kinetic ICL (**1**) generated reversibly with a half life of 3.5 h. The DOB ICL (**2**) is formed reversibly as well, albeit with a greater half life (10 h). Another intriguing fact about both the ICLs is their site-selective formation. The cross-links are generated predominantly with an A opposite the 3′-adjacent nucleotide.[8,9] Although lifetimes of both the ICLs are long enough to cause disruptions in cellular activities, their unstable nature renders their biochemical fates difficult to investigate. Likewise, the rationale behind the site-selective formation also remains unknown.

Scheme 1: ICL formation by DOB and C4-AP



In order to probe the biochemical fate and site-selective formation of the DOB and C4-AP kinetic ICLs, I synthesized stable mimics. Among the commonly used cross-link synthesis strategies, one involves direct treatment of the duplex with a cross-linking agent. This method often results in mixtures of monoadducts and unstable ICLs.[26] Another drawback of this method are the low yields of isolated ICLs.[27] An alternative

strategy is the use of post-oligomerization modification where a masked ICL precursor is incorporated into one of the strands by solid phase synthesis.[28] The precursor is unmasked after duplex formation leading to cross-link formation with a nucleotide from the opposing strand. A similar approach involves incorporation of masked ICL precursors in both strands and their conversion into an ICL following duplex formation.[29] However, this method is conceptually identical with the formation of C4-AP and DOB ICLs which were originally generated by reaction of an electrophile in one strand with a nucleobase from the other.[30,31] Hence, ICL mimics of **1** and **2** cannot be produced using this method.

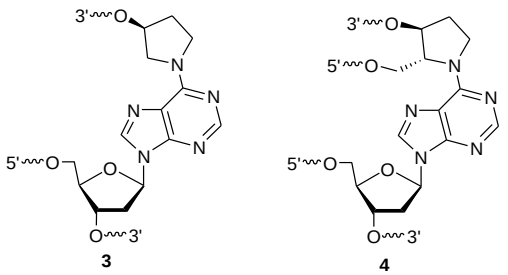


Figure 2. Structures of the synthetic mimics of DOB (**3**) and C4-AP kinetic ICL (**4**).

An additional drawback of this method is sequence restriction. Due to the employment of highly reactive monoadducts, the adjoining sequence is deliberately deprived of nucleophilic nucleobases to prevent unwanted cross-linking. The third approach is the direct synthesis of the cross-linked dinucleotide. The dinucleotides are then converted into phosphoramidites where one or more hydroxyls are masked with protecting groups. After the phosphoramidite is incorporated into DNA by solid phase synthesis, sequential removal of protecting groups and oligonucleotide synthesis using the free hydroxyl group are employed for the cross-link mimic synthesis.[32,33] Applicability of this method has been limited by either sequence symmetry or the lengths of the ICLs produced. We decided to upgrade the third method for the syntheses of mimics of **1** and **2**. The

hemiaminal linkages were identified as the source of instability in the actual ICLs (**1** and **2**) and were removed to stabilize the mimics. Therefore, the proposed mimics (**3** and **4**, Figure 2) were in lower oxidation states than those of the actual ICLs. We deduced that such minor changes in chemical structure would not significantly affect their structural aspects or interaction with repair proteins.

We report the synthesis of C4-AP and DOB ICL mimics using novel orthogonal deprotection strategies without any sequence restrictions. The strategies enabled synthesis of both mimics with oligonucleotide lengths appropriate for repair investigation. In an attempt to understand the biochemical fates of the ICLs, treatment of these mimics with UvrABC endonuclease is described. Finally, we hypothesized that site-selectivity of DOB and C4-AP ICL formation can originate as a result of attaining minimal deformation in the final ICL structure. We thus contemplated that chemical reagents, which react preferentially at the site of deformities would help us detect the stated structural distortion. Using the same line of reasoning we anticipated that we could employ thermal denaturation to probe the stability of the mimics. Therefore, we also outline the treatment of the mimics with chemical agents and thermal denaturation studies to uncover the reason behind site selective ICL formation.

2. Background

2.1. In vitro syntheses of interstrand cross-links (ICLs)

According to the central dogma of genetics, genetic information vital to the life processes flows from DNA to RNA via transcription. Another process responsible for existential maintenance of organisms is replication, which is carried out by DNA polymerases. These key processes involve strand separation of duplex DNA. Due to the

core role of DNA in two pathways for sustaining the life cycle of any cell including cancerous ones, DNA remains one of the favorite targets of drug molecules. One class of DNA lesions generated is the interstrand cross-link (ICL). ICLs are covalent linkages formed by reaction between nucleobases from opposing strands with a bis-electrophile.[15] Alternatively, they can also be formed upon reaction between nucleobases and DNA lesions (e.g. AP site,[6,7] C4-AP[8], DOB[9]). Typically, the bis-electrophiles are administered as a drug candidate for diseases like cancer. Many such bis-electrophiles like nitrogen mustards, cyanoethylnitrosourea (CENU), and mitomycin effectively fight diseases. Better understanding of the DNA biochemistry lead to discovery of bifunctional cross-linking agents that are not bis-electrophiles, such as psoralens. Ironically, bis-electrophiles such as malondialdehyde which are capable of cross-linking were discovered to be side products of the cell's own metabolism. Similarly, cross-links are reported to be produced from damaged nucleotides themselves.

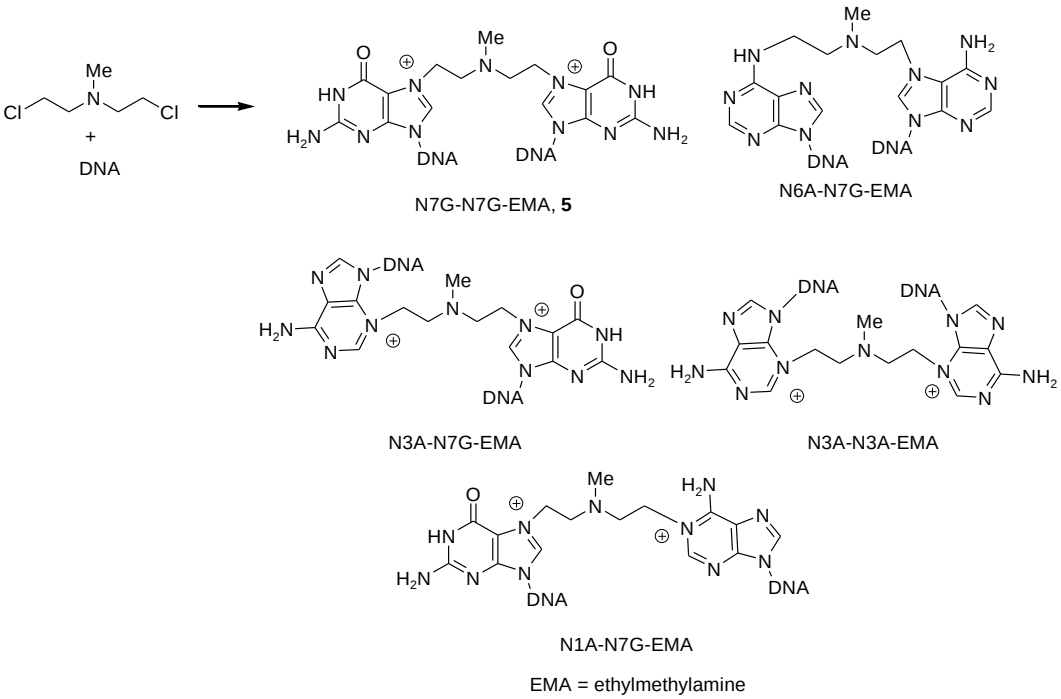
In order to comprehend the biochemical nature of damage caused by ICLs, it is necessary to synthesize ICLs in a pure and sequence specific manner. These synthesized ICLs can be studied in vitro using repair proteins or can be transfected into the cell by a plasmid. The structures of the synthesized ICLs are also studied in depth due to direct implication with repair. In this report we discuss the existing methods of ICL synthesis, their advantages, drawbacks, the effect of ICL formation on DNA structure and DNA repair pathway in bacterial cells.

2.1.1. ICL generation by treatment of chemical agents with DNA

2.1.1.1. ICLs generated from nitrogen mustard reagents

Nitrogen mustards (NM) are bifunctional agents widely used against leukemia, lymphoma, multiple myeloma, and ovarian carcinoma.[15] Although quite effective as anti-cancer agents, classical NM drugs are non-selective in nature and lead to undesired cytotoxicity in non-cancerous cells.[34] One such agent, N,N-bis(2-chloroethyl)methylamine (mechlorethamine) predominantly reacted with N7s of deoxyguanosine (G) to generate N-(2-[N7-guanyl]ethyl)-N-(2-[N7-guanyl]ethyl)

Scheme 2. Cross-link chemical structures generated by reaction of mechlorethamine with DNA



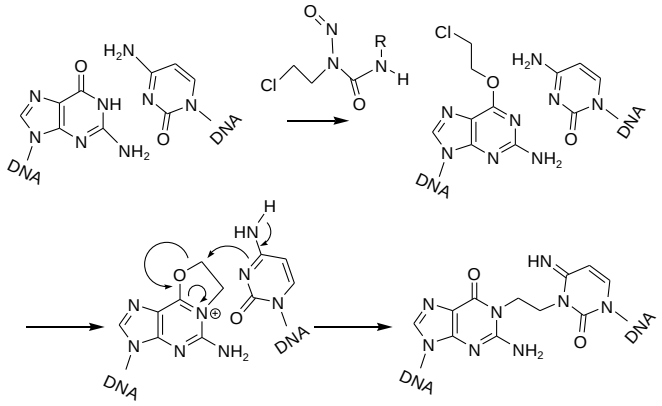
methylamine (N7G-N7G-EMA, **5**). Additionally, mechlorethamine ICLs were generated between N3 of deoxyadenosine (dA) and N7 of G (N3A-N7G-EMA), N1 of A and N7 of G (N1A-N7G-EMA), N6 of A and N7 of G (N6A-N7G-EMA), and between N3s of A (N3A-N3A-EMA).[35] N7G-N7G-EMA ICL was predominantly formed in 5'-GNC

sequence.[1] N1A-N7G-EMA, and N6A-N7G-EMA ICLs were formed in 5'-GGT sequence, while N3A-N7G-EMA ICLs are formed in 5'-GT sequences.[35] No sequence selectivity was found for N3A-N3A-EMA in the literature. The alkylation of bases makes them labile to depurination with half lives ranging from hours to days.[26,36] The ICL yield is typically low and varies from 1-6%.[36] Generation of a mixture of products combined with low yield and instability render any repair analysis for these ICLs difficult.

2.1.1.2. ICL generated from Chloroethylnitrosoureas (CENU)

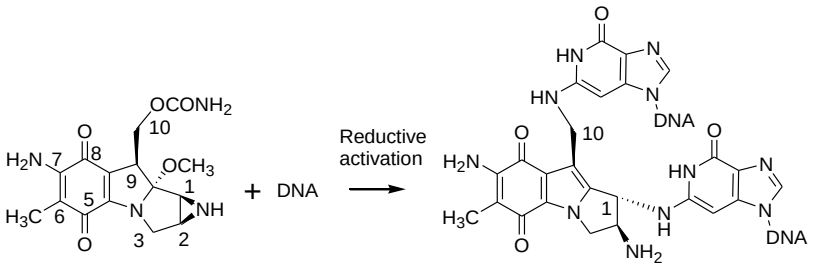
Chloroethylnitrosoureas (CENU) are commonly used to treat brain tumors due to their robust potential in traversing the blood brain barrier.[37] CENU ICL (**6**) did not show any sequence selectivity and formed the ICL between N3 of C and N1 of the opposing G (Scheme 3).[27] In contrast to NM ICLs, which were formed through either the major or minor groove, CENU ICL was formed at the hydrogen bonding face. However, the ICL yield upon direct treatment of DNA with CENU reagent was low (0.4 – 3.7%), making their structural characterization and repair investigation difficult.[27]

Scheme 3. DNA ICL formation by treatment with chlorethylnitrosourea (CENU)



2.1.1.3. ICL generated from mitomycin C

Scheme 4. ICL formation by mitomycin C



N2G-N2G-MC, **7**

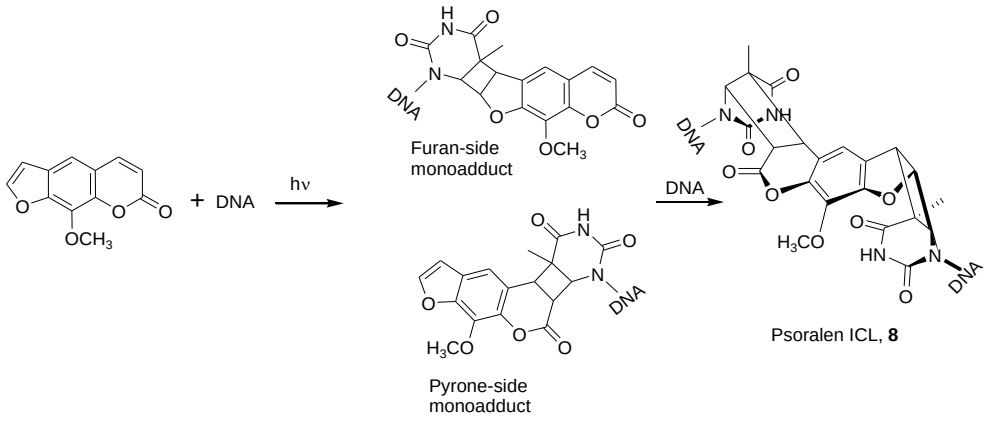
Originally isolated as an antibiotic metabolite from *Streptomyces caespitosus*, mitomycin C has been effectively used in chemotherapeutic treatment of various forms of gastrointestinal tumors.[16] Upon reduction, mitomycin C is converted to a bis-electrophile capable of cross-linking with N2 of Gs on opposing strands (Scheme 4). The minor groove ICL (**7**) was formed in a 5'-(GC) sequence.[38] In addition to forming of monoadducts and intrastrand cross-links,[38,39] the ICL yield varies from 5 – 50% when a single 5'-(GC) was present in a duplex[40,41] The cross-link was generally stable, but decomposed upon treatment with hot piperidine or treatment with reducing agents like $Na_2S_2O_4$ and dithiothreitol (DTT).[41]

2.1.1.4. ICL generated by psoralen

Psoralens belong to the family of fucomycin natural products and are planar compounds isolated from a number of diverse sources ranging from plant seeds to bacterial broth.[42,43] They intercalate into DNA and upon UV irradiation form ICL **8** in a 5'-(AT) sequence with C5-C6 double bond of T (Scheme 5).[44] This major groove ICL forms in a modest yield (5%).[45] Due to their ability to form ICLs in a light controlled manner, they are used against skin diseases such as psoriasis.[17] The side products of

psoralen compounds reacting with DNA are the pyrone-side and furan-side monoadducts while no intrastrand cross-links have been reported.[44]

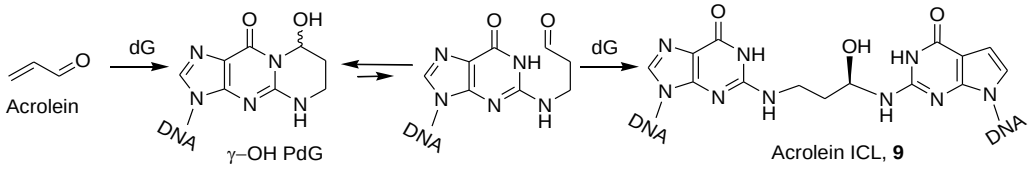
Scheme 5. Reaction of 8-methoxypsoralen with DNA to generate monoadducts and ICL



2.1.1.5. ICL generated by aldehydes

Aldehydes like acrolein and malondialdehyde (MDA) contributes towards minor groove ICL generation in DNA. Acrolein has endogenous (lipid peroxidation[46]) as well as exogenous origins (cigarette combustion[47]), while MDA is generated as an unstable side product of lipid peroxidation.[48] Both acrolein and MDA are mutagenic in mammalian and bacterial cell assays.[49–52] The bis-electrophiles generate ICLs by reacting with N2 of G in a 5'-(CG) sequence (Scheme 6

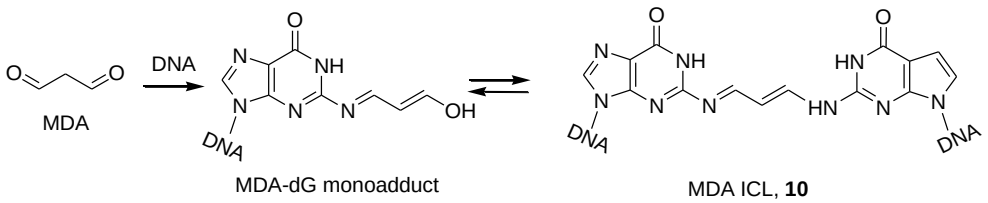
Scheme 6. ICL generation by acrolein



and Scheme 7). Due to inherent instability of the hemiaminal bond formed, 20% of acrolein ICL **9** reverts to un-cross-linked duplex in 16 h at 25 °C and pH 7.0.[53] Under

conditions that disfavor duplex formation (absence of buffer), the reversal was faster and complete within 1 h.[53] When DNA was directly treated with acrolein the ICL yield remained low[54] and was accompanied by γ-hydroxy-1,N2-propanodeoxyguanosine (γ-OH-PdG) as the major monoadducts.[55] The monoadduct was also reported to generate intrastrand cross-link with 5'-adjacent A, T and Gs.[56] In contrast, no intrastrand cross-link has been reported for MDA monoadduct. MDA ICL **10** was recovered in only 0.05% yield.[57]

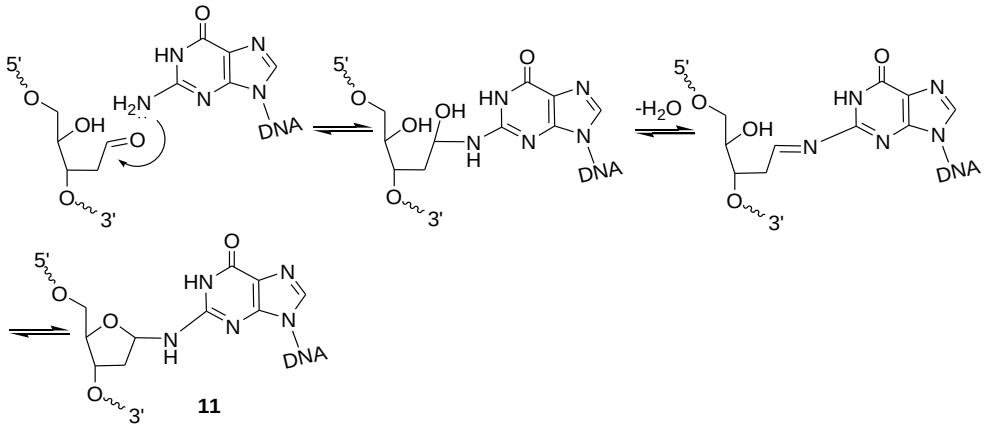
Scheme 7. ICL generation by MDA



2.1.1.6. ICLs generated by various abasic sites

2.1.1.6.1. ICLs generated by an AP site

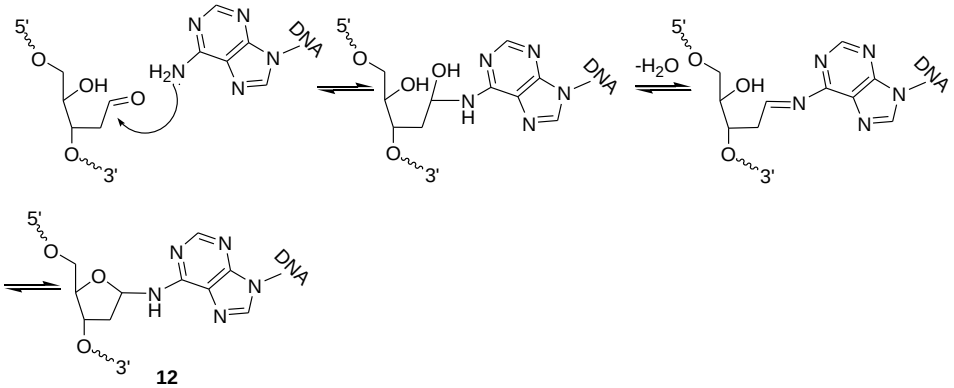
Scheme 8. ICL formation between abasic site and dG



AP sites are DNA lesions resulting from enzymatic or non-enzymatic hydrolysis of a glycosidic bond with approximately 10000 sites generated per cell every day.[58,59] AP

exists in equilibrium between a ring opened carbonyl and a ring closed hemiacetal structure. Recently, Gates and coworkers reported that AP cross-links with G and A.[6,7] Generation of AP-G ICL **11** occurred in 18% yield and involved reaction with guanosine opposite the 5'-adjacent nucleotide of AP. The N2 amine of G participated in ICL formation (Scheme 8). This minor groove ICL formation was unaffected by pH alteration (3 – 10). Although methoxyamine presence in the beginning of the assay severely inhibited cross-link formation, post-formation treatment of the ICL with methoxyamine

Scheme 9. ICL formation between abasic site and dA

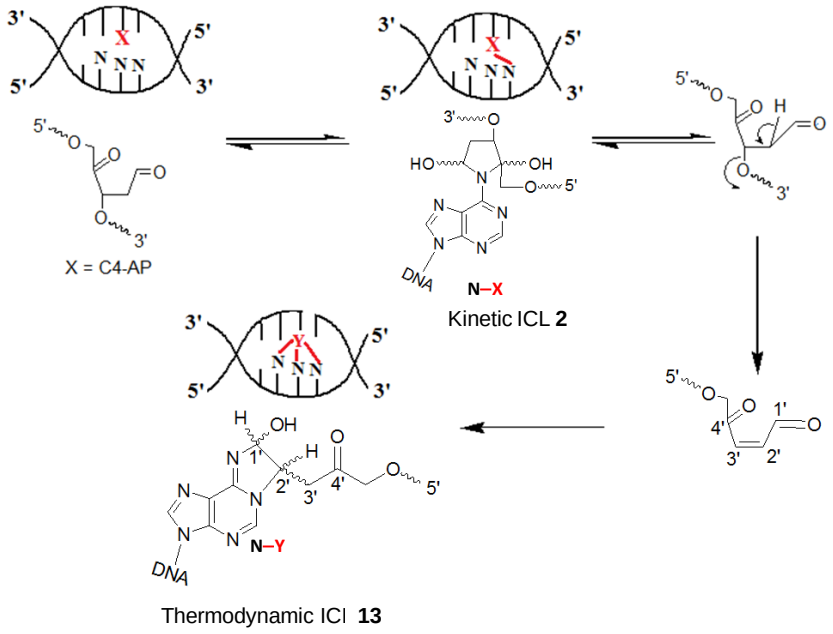


or hydrazine (aldehyde trapping reagents) did not decrease the yield.[6] In contrast to G, the participating A was exclusively the one opposite the 3'-adjacent nucleotide of the AP with the N6 amine acting as a nucleophile (Scheme 9). The AP-A ICL **12** was formed in 13-71% yield depending on the sequence and was stable within the pH range 3 – 10 or mild piperidine workup. In contrast to the G-AP ICL, this major groove ICL decomposed when heated at 90 °C or upon treatment with aldehyde trapping methoxyamine, signifying a possible difference in equilibrium chemical structure.[7]

2.1.1.6.2. ICL generated by C4'-oxidized abasic site (C4-AP)

C4-AP is generated from C4′ oxidation of the deoxyribose sugar by exposure to γ–irradiation,[24] radiomimetic agents such as the bleomycin family of glycopeptides,[25] and various enediynes (neocarzinostatin[25], esperamicin[60]). Bleomycin coordinates with Fe (II) and oxygen in blood plasma to form an activated ternary complex that has the properties of iron-superoxide complexes. The activated species binds in the minor groove of DNA. The O-O bond then undergoes heterolytic cleavage and forms a (O)Fe(III) species

Scheme 10. Kinetic and thermodynamic ICL formation by C4-AP



responsible for C4′ hydrogen abstraction. The damage is sequence selective, and occurs primarily on the pyrimidine residues 3′-to a guanosine: 5′-(GC) and 5′-(GT).[25] Abstraction of the C4′-hydrogen is postulated to lead to two pathways. In the oxygen dependent pathway, it produces 3′-phosphoglycolate and base propenal. In an oxygen independent pathway it forms an alkali labile 4′-hydroxylated abasic site (C4-AP).

Neocarzinostatin (NCS) also binds in the minor groove of DNA. Nucleophilic addition of a thiol to NCS leads to spontaneous rearrangement of its enediyne core to a reactive carbon centered-diradical. The diradical abstracts C4′-H and generates C4-AP in an oxygen dependent manner.[25]

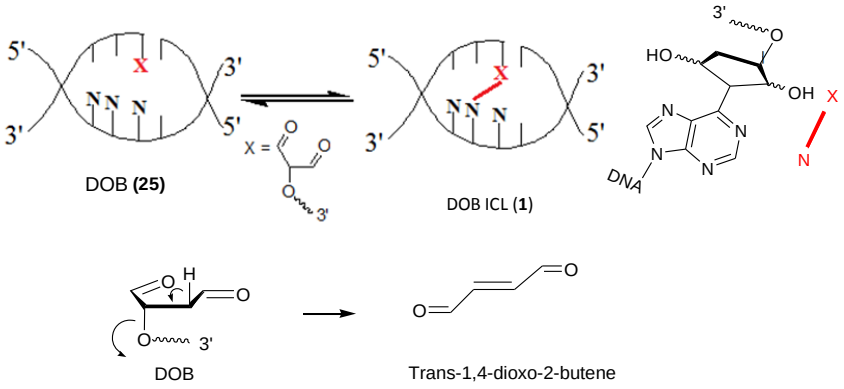
Regulus *et al.* reported that C4-AP generated from ionizing radiation or bleomycin treatment of calf thymus DNA forms an adduct with C, although the exact mechanism of its formation was unclear.[61] Recently, it was demonstrated that C4-AP generates a high (**2**) and a low (**13**) molecular weight ICL. Reversible adduct formation with A opposite the 3'-adjacent nucleotide of C4-AP provided the high molecular weight ICL (**2**) in yields ranging from 16 – 37%. Similar to AP-A ICL, the N6 amino group of A was involved in ICL formation. The ICL was unstable with a half life of 3.1 h and reverted back to C4-AP. β-Elimination resulting in a strand scission and generation of an α,β-unsaturated aldehyde effected further reaction with dA and led to irreversible formation of low molecular weight ICL **13**. ICL **2** was formed 10 times faster ($k_1 = 3.0 \pm 0.1 \times 10^{-5}$ s^{-1}) than **13** ($k_2 = 3.4 \pm 0.3 \times 10^{-6}$ s^{-1}), establishing **2** as the kinetic product. Importantly, the site selective formation of the **2** with dA opposite the 3'-adjacent nucleotide could not be justified.[31]

2.1.1.6.3. ICL generated by dioxobutane (DOB)

C5′-oxidation of the deoxyribose sugar originates from the assault of a variety of DNA-damaging agents including enediynes (neocarzinostatin, calicheamicin),[25,62–64] manganese porphyrins,[65] and γ–radiolysis.[66] In the case of enediynes, the same carbon centered diradical generated acts as the hydrogen atom scavenger. One of the major

products formed from such oxidation was the 5′-(2-phosphoryl-1,4-dioxobutane), DOB.[62] Generation of the dioxobutane species was observed during treatment of a short oligonucleotide with neocarzinostatin.[67] The DOB lesion was detected in aerobic γ–radiolysis of thymidine as well.[66] The DOB lesion eventually eliminates to form trans-1,4-dioxo-2-butene (Scheme 11).[63] The latter pathway is biologically important because cis-1,4-dioxo-2-butene, a furan metabolite, is a carcinogen in rodents.[68]

Scheme 11. ICL generation by DOB lesion



Participation of dioxobutane in interstrand cross-link formation preferentially with a A opposite a 3′-adjacent T of the damaged nucleotide was first reported by Guan.[9] When the cross-link (**1**) was characterized by ESI-MS, the adduct formation was found to involve the N6 amino group of A. The cross-link yield varied between 23-32%. Generation of the cross-link was reversible with a 10.1 h half life, while both the formation (k_1 = 2.6 × 10^{-5} $s^{-}1$) and decay (k_{-1} = 1.9 × 10^{-5} s^{-1}) of ICL were first order processes.

The equilibrium constant (K_{eq} = 1.4) of the ICL formation between A and DOB indicated that adduct formation was not significantly stabilized.[9] Intriguingly, there was no ICL formation with dC despite a comparable equilibrium constant (K_{eq} = 1.2) for

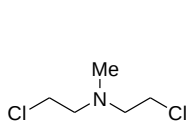
reaction with 1,4-butanedial, a DOB analogue.[9] In fact, the rate of the forward reaction of 1,4-butanedial with C ($k_{dC} = 9.3 \times 10^{-5}\ s^{-1}$) was almost 10 times higher than that with A ($k_{dC} = 1.2 \times 10^{-5}\ s^{-1}$). The distances between the C1'-aldehyde of DOB and the exocyclic amines of A and C from the opposing strand were compared by Guan using a Spartan model. When bases were positioned opposite the 3'-adjacent nucleotide of DOB, the exocyclic amine of the former was closer by almost 1 Å (7.03 Å for A vs 8.04 Å for dC). However, whether the larger distance predicted by the model was the determining factor behind ICL formation remains debatable. There was modest ICL formation (yield 3.2%) with a dA placed opposite the lesion even though its exocyclic amine was within 7.39 Å of the DOB aldehydic carbon.

2.1.2. Synthesis of Interstrand cross-links using post-oligomerization modification.

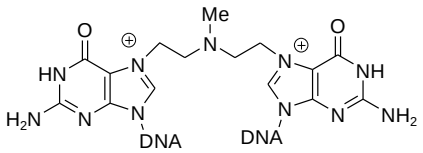
The origin of ICLs can be traced to a vast spectrum of DNA assaults ranging from therapeutic treatment with exogenous chemical agents and radiolysis, to endogenous metabolites, to deoxyribose sugar oxidation products. Due to the threat posed by ICLs to cellular pathways, organisms evolved to develop specific repair capabilities to deal with them. In the absence of such repair pathways, individuals can suffer from disorders such as Fanconi anemia.[69] In this cancer prone genetic disease, cellular pathways responsible for surveillance and removal of interstrand cross-linked DNA are crippled, making the patients vulnerable to cross-linking agents.[69] However, multiple aspects of such repair pathways still remain unexplored, making the associated research highly coveted in terms of drug designing and DNA damage diagnosis. A major limitation of such investigations is the dearth of pure cross-linked DNA due to extremely low isolation yields, generation of mixtures of products, and instability of the cross-links. Bioorganic chemists have

responded with two strategies to generate pure ICLs in high yields. The first one is based on post-oligomerization modification, where a nucleotide containing the masked monoadduct of a bifunctional agent is introduced into one or both strands of DNA using solid phase synthesis. After the two strands are annealed, the protecting groups are removed and the cross-link generated. In the second approach involving orthogonal deprotection strategy, a dinucleotide containing the stabilized cross-link is synthesized. The hydroxyls of the dinucleotide are masked using various protecting groups. Upon incorporation into the DNA using solid phase synthesis, the hydroxyls are unmasked sequentially and orthogonally. Finally, complete DNA strands are synthesized using the free hydroxyls. Both methods have their advantages and limitations as illustrated below.

2.1.2.1. Nitrogen Mustard ICL



EMA = N,N-bis(2-chloroethyl)methylamine, a nitrogen mustard reagent

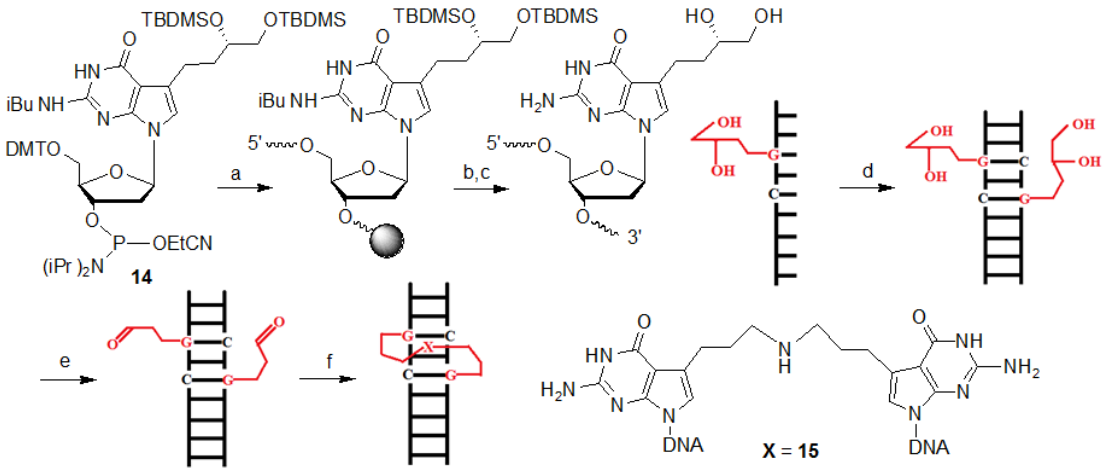


N7G-N7G-EMA adduct, **5**

Attempts to synthesize nitrogen mustard (NM) ICLs have been limited to dG-NM ICL adducts. To bypass the susceptibility of the N7G-N7G-NM adduct (e.g. **5**) to glycosidic bond hydrolysis, Schärer and coworkers employed isosteric 7-deaza analogues of G (Scheme 12).[29,70,71] The phosphoramidite containing the TBDMS-protected dialcohol (**14**) was incorporated into DNA using solid phase oligonucleotide synthesis. After the nucleobase and alcohol protections were removed, the single strand with the unprotected diol containing dG analogue was annealed with a complementary strand

having the same modification. The G-NM-G ICL analogue (**15**) was then synthesized in 25-75% yield after oxidation of the duplex, followed by reductive amination.[70]

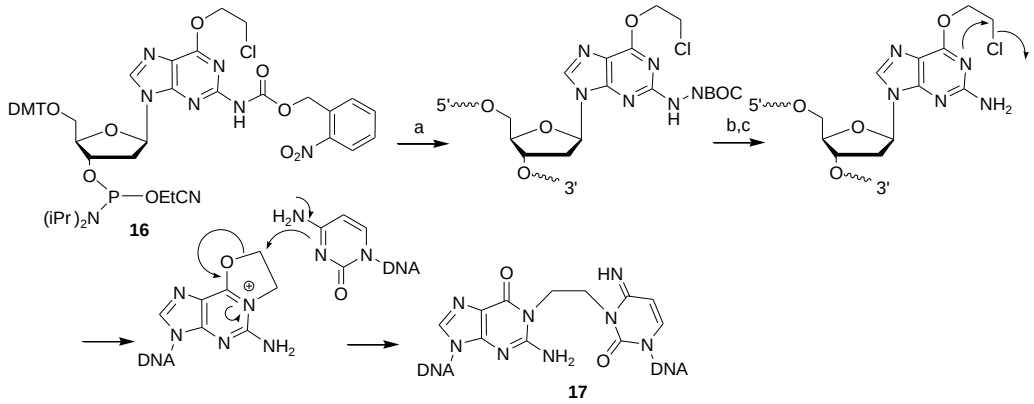
Scheme 12. Synthesis of nitrogen mustard ICL analogue by post-oligomerization modification



Key: (a) Solid phase oligonucleotide synthesis; (b) NH_3, 50° C, 16 h; (c) TEA.3HF, 40° C, overnight; (d) annealing; (e) $NaIO_4$, pH 5.4; (f) NH_4OAc, $NaCNBH_3$.

2.1.2.2. ICL made by chloroethylnitrosourea

Scheme 13. Synthesis of chloroethylnitrosourea ICL



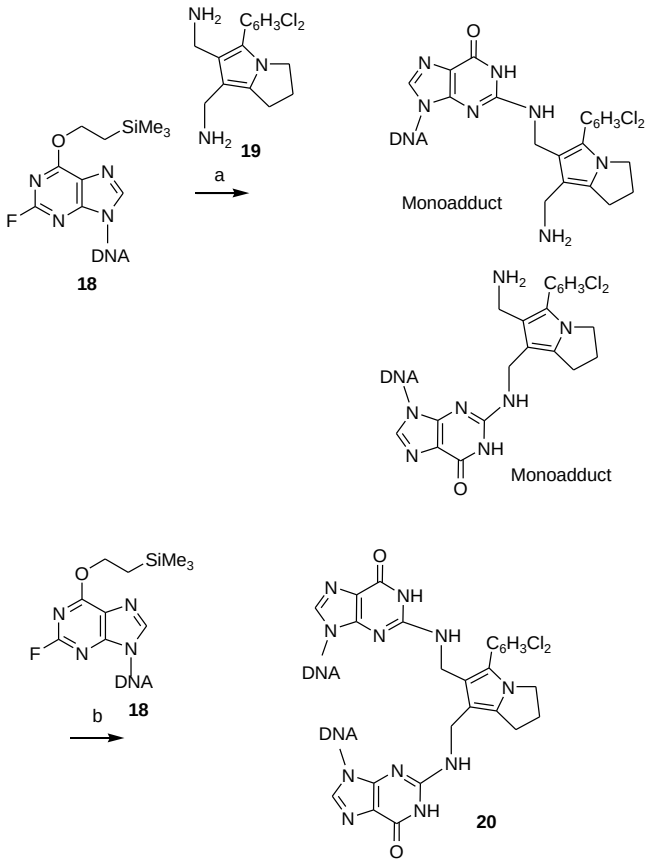
Key: a) Solid phase synthesis; b) hν 365 nm, 3 min, room temperature; c) 17 h incubation, 37 °C.

A mimic of chloroethylnitrosourea (CENU) ICL between the N3 of C and the N1 of G was synthesized using a photocaged nucleobase analogue (Scheme 13).[72] In this

approach, O6-(2-chloroethyl) and N2-ortho-nitrobenzyloxycarbonyl (NBOC) protected G phosphoramidite (**16**) was incorporated into DNA using solid phase synthesis. The electron withdrawing nature of NBOC reduced the nucleophilicity of the N1 positon preventing its nucleophilic attack on the chloroethyl group. After annealing with the complementary strand, the NBOC group was deprotected by irradiation at 365 nm, and N1 displaced the chloride. Attack by N3 of the opposing C freed the O6 of G generating the ICL (**17**). The ICL was synthesized in 37% yield. Interestingly, the presence of A, T or G opposite NBOC modified G resulted in ICL generation as well.

2.1.2.3. Synthesis of interstrand cross-links of bifunctional pyrroles

Scheme 14. Synthesis of ICLs between dG and bifunctional pyrroles



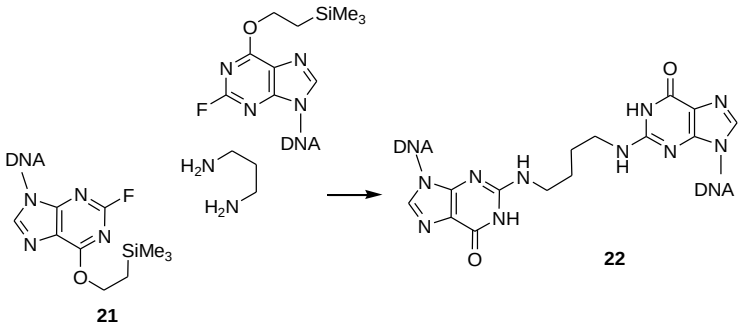
Key: (a) 20 °C, 4 days; (b) 20 °C, 8 days.

Synthesis of a mimic of an ICL between G and 2,3-dihydro-pyrrolizine bis(isopropylcarbamate) (IPP) was accomplished in a reaction in which the traditional roles of the nucleobase and the cross-linking reagent were inverted (Scheme 14).[73] The nucleobase was a fluoropurine that served as the electrophile, while the cross-linking bifunctional pyrrole functioned as the nucleophile. In this synthesis, O6-protected fluorodeoxyinosine incorporated in ssDNA (**18**) was treated with diamine **19**, an IPP analogue. This reaction resulted in formation of monoadducts in 40% yield that were then treated with **18** to obtain cross-link **20**.

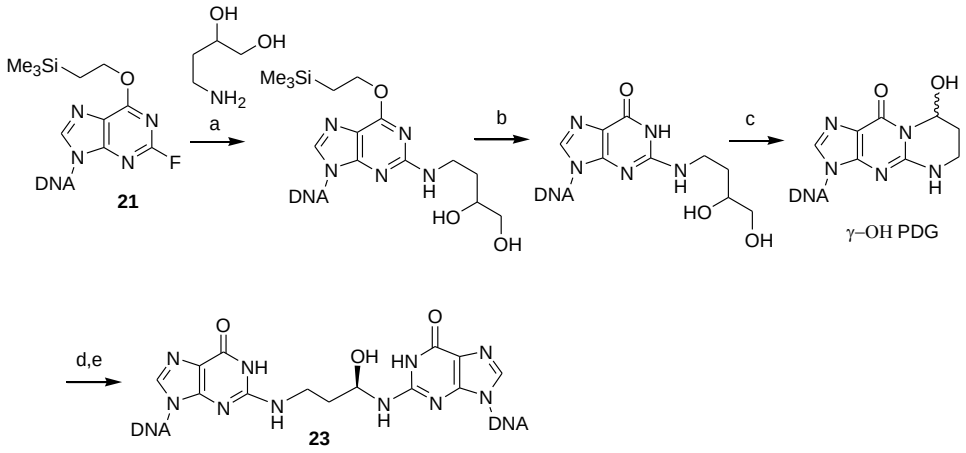
2.1.2.4. Synthesis of interstrand cross-links from aldehydes

Stable cross-link adducts between DNA and bis-electrophilic aldehydes were synthesized by two different post-oligomerization modifications. In the first, the saturated N2-N2-propano analogue (**22**) of malondialdehyde ICL was synthesized in a 5'-(CG) motif by Harris and co-workers using nucleophilic attack of the alkyl amine on fluoronucleoside **21** (Scheme 15).[74] In the second method (Scheme 16), 1,N^2-γ-hydroxypropano-dG (γ-OH-PDG), the monoadduct between acrolein and G was synthesized from the O6-protected fluoronucleoside **21** in a stepwise manner. After reaction of **21** with 4-amino-1,2-butanediol, subsequent removal of the O6-protection and oxidation of the vicinal diol produced γ-OH-PDG. After annealing with its complement, its reaction with an opposing strand dG in a 5'-CpG sequence for 7 days at 25 °C produced ICL **23** with 50% yield.[29] The cross-link **23** showed the same stability as reported for the isolated ICL from reaction between DNA and acrolein.[53]

Scheme 15. Synthesis of saturated N2-N2-propano analogue of malondialdehyde ICL



Scheme 16. Synthesis of acrolein-dG ICL

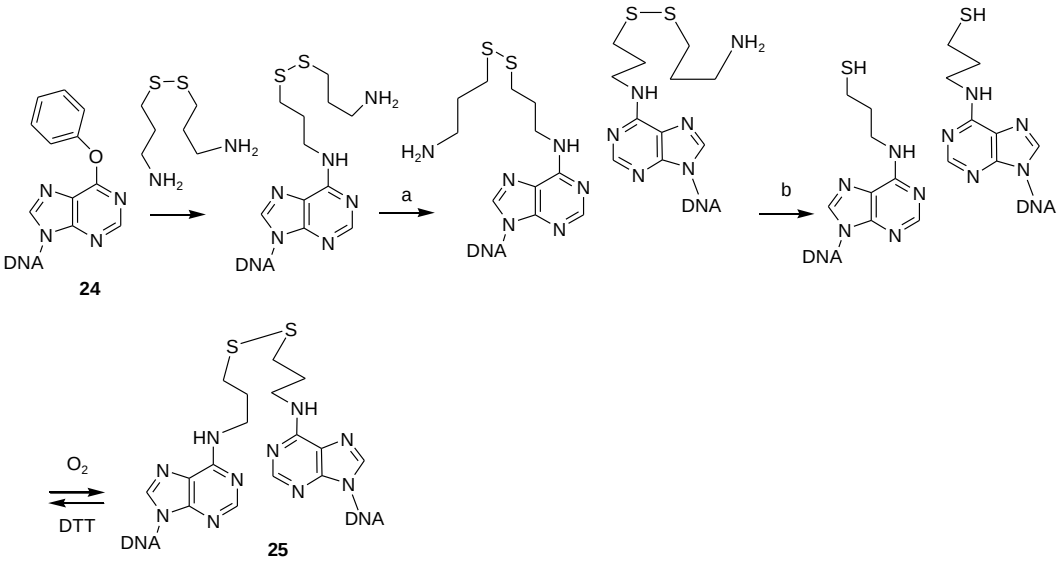


Key: (a) DMSO, $iPrEt_2N$, 65°C; (b) 0.2 M HCl; (c) $NaIO_4$; (d) Annealing; (e) 25°C, 7 days.

2.1.2.5. Synthesis of oligonucleotides containing disulfide interstrand cross-links

Utilizing a nucleophilic attack of 3,3'-dithio bis(propylamine) on convertible nucleoside O6-phenyl-2'-deoxyinosine (**24**) the alkane disulfide tethers were attached to the N6-amino group of dA (Scheme 17).[75] In a duplex containing two N6-alkane-disulfide-tethered dAs in a 5'-(AT) sequence, incubation with dithiothreitol (DTT) converted it to a N6-alkane. Upon removal of DTT from the solution, the sulfides from opposing strands were spontaneously linked to generate a site-specific alkane-tethered disulfide ICL (**25**) in quantitative yield. Introduction of DTT led to reversal of the cross-linked duplex to un-cross-linked N6-alkane sulfide.

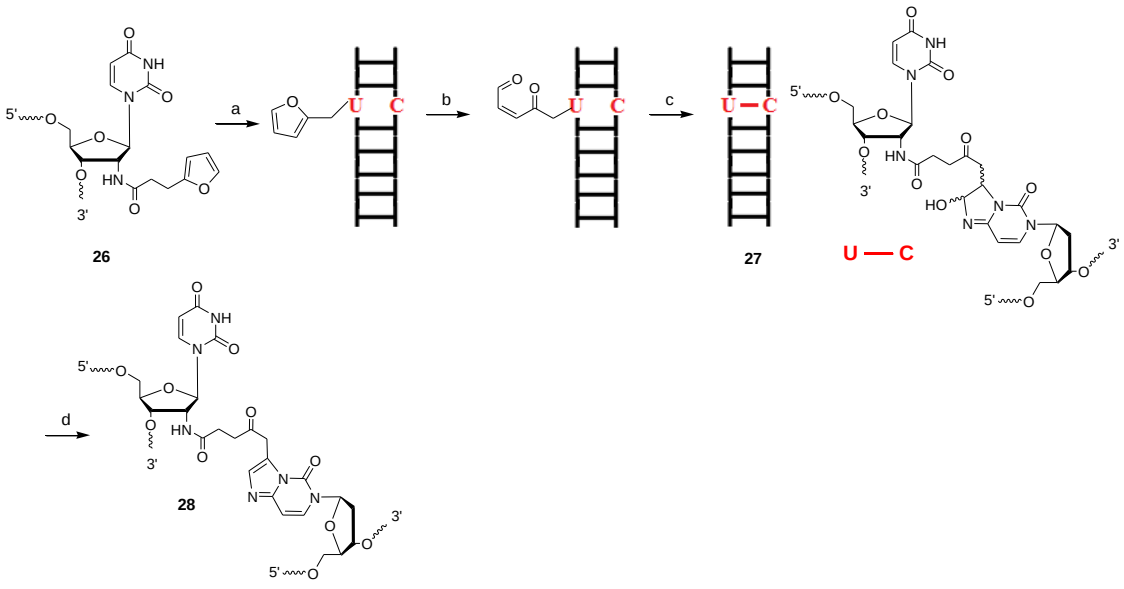
Scheme 17. Synthesis of a site-specific disulfide ICL



Key: (a) Annealing with complementary strand; (b) DTT.

2.1.2.6. Cytidine selective cross-link generation from 2'-O-furan modified dU and 2-amino-6-vinylpurine

Scheme 18. dC selective cross-linking by 2'-O-furan modified U

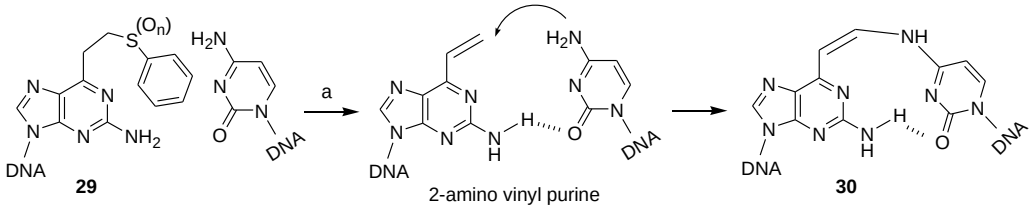


Key: (a) Annealing with complementary strand; (b) N-bromosuccinimide; (c) incubation, 15 min, 25 °C; (d) 1-3 days, 50 °C.

In this strategy exploiting the natural cytotoxicity of furan, the 2'-O position of uridine was modified with a masked furan as a γ-keto enal (**26**, Scheme 18).[30] When a duplex containing masked 2'-O-furan-U in one strand was oxidized, cross-linking occurred with an opposing dC producing ICL **27** (Scheme 18) with an average 20% yield. The unstable product containing a hemiacetal linkage was converted into stable **28** by aromatization upon heating at 50°C.[30]

In another instance of site selective cross-link generation with C, a phenylsulfoxide precursor of 2-amino-6-vinylpurine (AVP, **29**) was incorporated into DNA via solid phase synthesis (Scheme 19).[76] After the strand containing the precursor was annealed with a complement, in situ elimination of phenylsulfoxide generated 2-amino-6-vinylpurine that cross-linked with the opposing dC, producing **30** in 82% yield after 24 h.

Scheme 19. Generation of C selective cross-linking using 6-amino2-vinylpurine

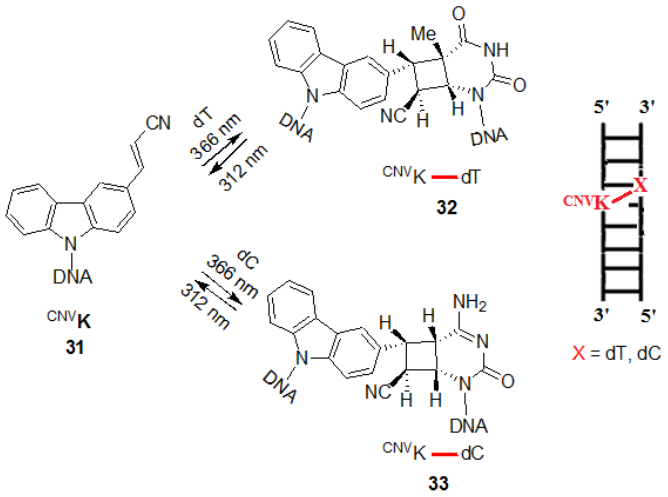


Key: (a) Incubation in 50 mM 2-(N-morpholino)ethanesulfonic acid (MES) buffer, pH 5.0, 100 mM NaCl.

2.1.2.7. Photoinduced cross-linking reactions

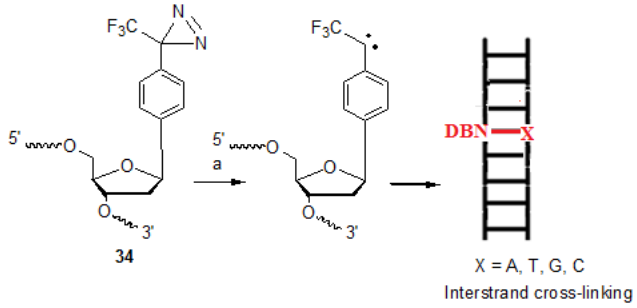
Irradiation of 3-cyanovinylcarbazole nucleoside (^{CNV}K, **31**) led to efficient ICL formation with T (**32**) or C (**33**) opposite the 5'-adjacent nucleotide (Scheme 20).[77,78] Quantitative ICL generation with C and T occurred within 5 s and 60 s of 366 nm light exposure, respectively. Irradiation at 312 nm reverted the ICL to non-cross-linked duplex.

Scheme 20. Interstrand cross-linking by ^{CNV}K

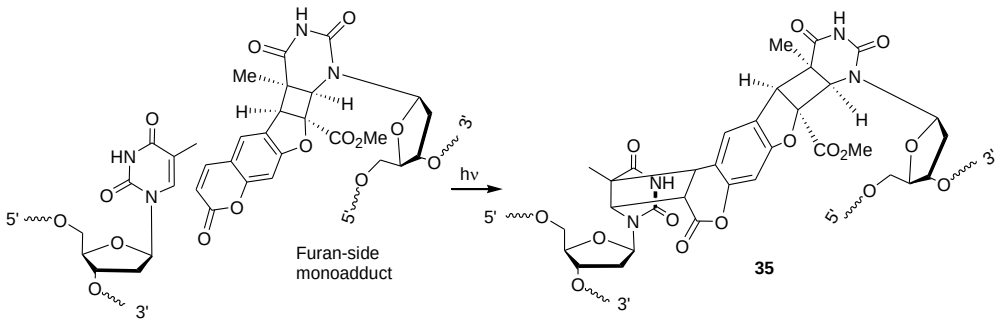


In an example by Chuan He, a highly reactive carbene intermediate generated from a diazirine-based nucleotide analogue (**34**) formed an interstrand cross-link with the opposite nucleotide irrespective of its identity (Scheme 21).[35] The cross-linking reaction was complete after 10 min of 350-360 nm irradiation, but the chemical structures of the cross-links were not characterized.

Scheme 21. ICL generation from diazirine-based nucleotide analogue

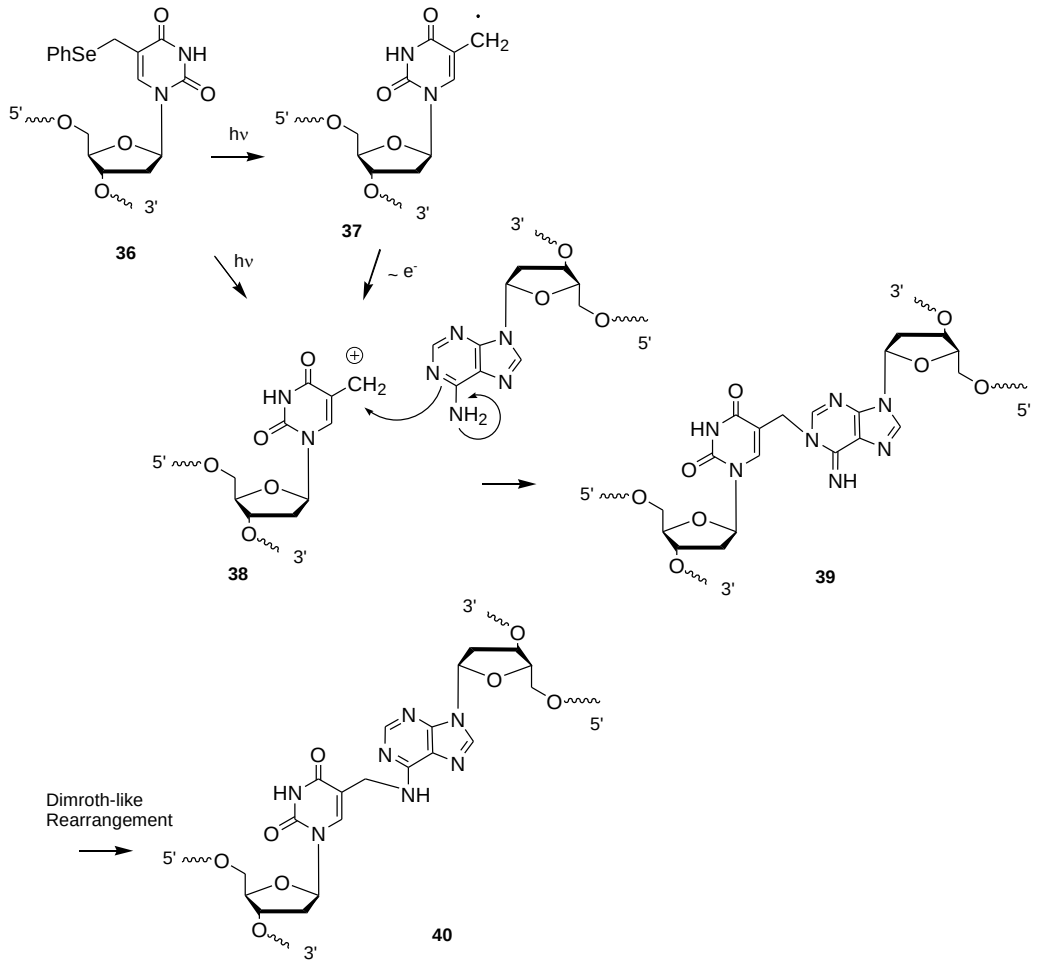


Scheme 22. Generation of site selective Psoralen ICL



Site-selective generation of a psoralen ICL was achieved by Koberts and co-workers using a furan-side monoadduct of the bi-functional photocross-linking agent

Scheme 23. ICL generation by phenyl selenide modified thymidine



incorporated in one of the strands (Scheme 22). When annealed with a complement with the monoadduct residing in a 5'-TA local sequence, irradiation with 366 nm UV light yielded the cross-link (**35**) in 84% yield.[79]

Photolysis and subsequent reactions of **36** have served as a model for the reactivity of 5-(2′-deoxyuridinyl)methyl radical **37** that was generated upon γ-radiolysis of DNA (Scheme 23).[80] Recently, the Greenberg laboratory demonstrated the cross-linking ability of a phenylselenyl modified thymidine **36** with A upon photolysis.[81] Single

electron oxidation of **37** and/or heterolysis of the C-Se bond in **36** upon photolysis generated carbocation **38**, which reacted with the N1 of A to generate ICL **69** in 25% yield.[82] Utilizing an associative (Dimroth-like) rearrangement, **39** isomerized to **40.**[83]

2.1.3. Use of orthogonal deprotection strategy for synthesizing ICL

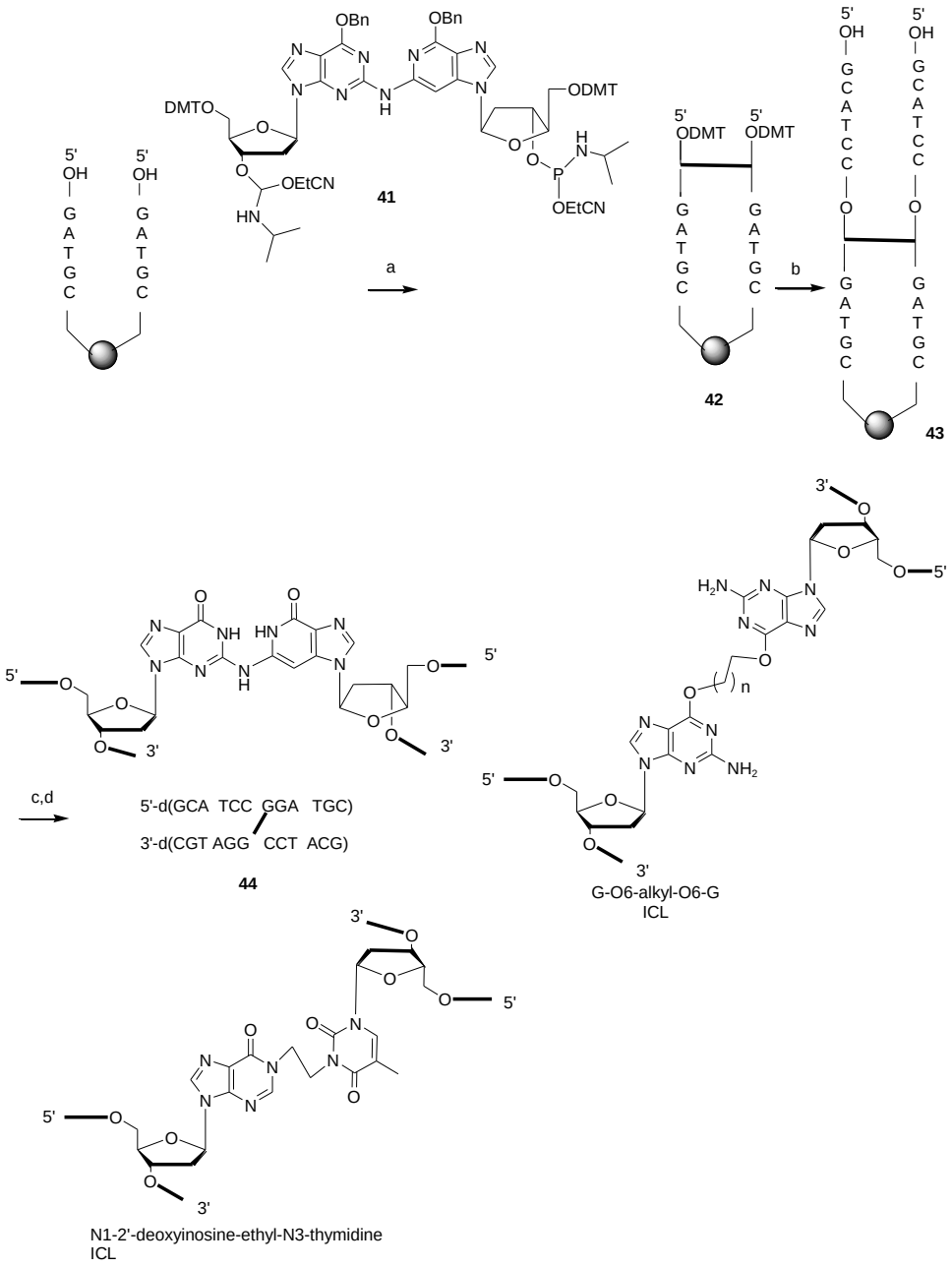
2.1.3.1. Synthesis of nitrous acid ICL

The first example of this approach was demonstrated by the application of cross-linked bis-phosphoramidite (**41**) for the synthesis of a nitrous acid cross-link (Scheme 24). More heavily loaded solid support was employed where **41** simultaneously coupled with two growing oligomers yielding **42**. Continued oligomeric synthesis generated support bound **43** that was subjected to ammonia deprotection and PAGE purification.[84] Finally, the O6-benzyl groups were removed by transfer hydrogenolysis on isolated DNA to yield palindromic ICL **44** in 10% overall yield. A similar strategy has been used to synthesize G-O6-alkyl-O6-G and N1-2′-deoxyinosine-ethyl-N3-thymidine ICL (Scheme 23).[85,86]

2.1.3.2. Synthesis of N^4C-ethyl-N^4C cross-linked DNA

Synthesis of a N4C-ethyl-N4C cross-link, a proposed structural mimic of the mechloroethamine adduct with a C-C mismatch was achieved using phosphoramidite 45 (Scheme 25).[87] The phosphoramidite was coupled with an oligomer on a solid support to yield 46, that upon continued synthesis and end-capping, generated 47. After removal of the TBDMS group, the synthesis of the full length support-bound cross-link (48) was completed by coupling 3'-cyanoethyl “reverse” phosphoramidites. Ammonolysis yielded

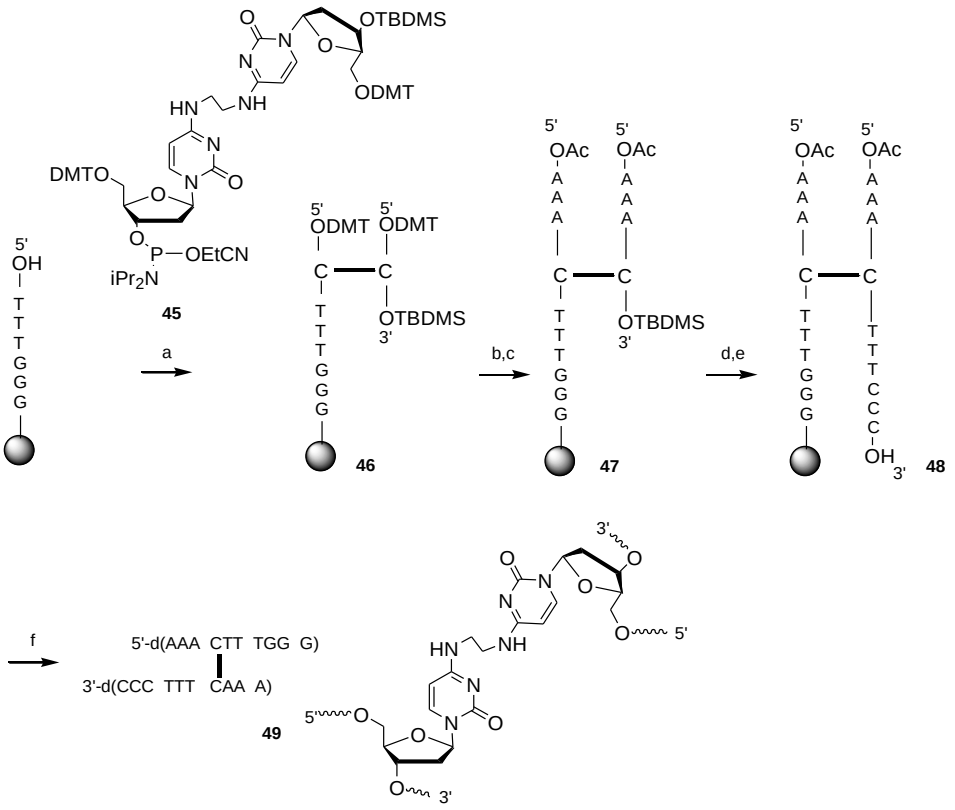
Scheme 24. Synthesis of nitrous acid ICL



Key: (a) Coupling of **41** during Solid phase oligonucleotide synthesis; (b) oligomer synthesis; (c) ammonolysis and purification; (d) hydrogenolysis.

ICL **49** containing 3'-overhangs that could be used to generate longer ICLs via enzymatic ligation.[32]

Scheme 25. Synthesis of N4C-ethyl-N4C cross-linked DNA



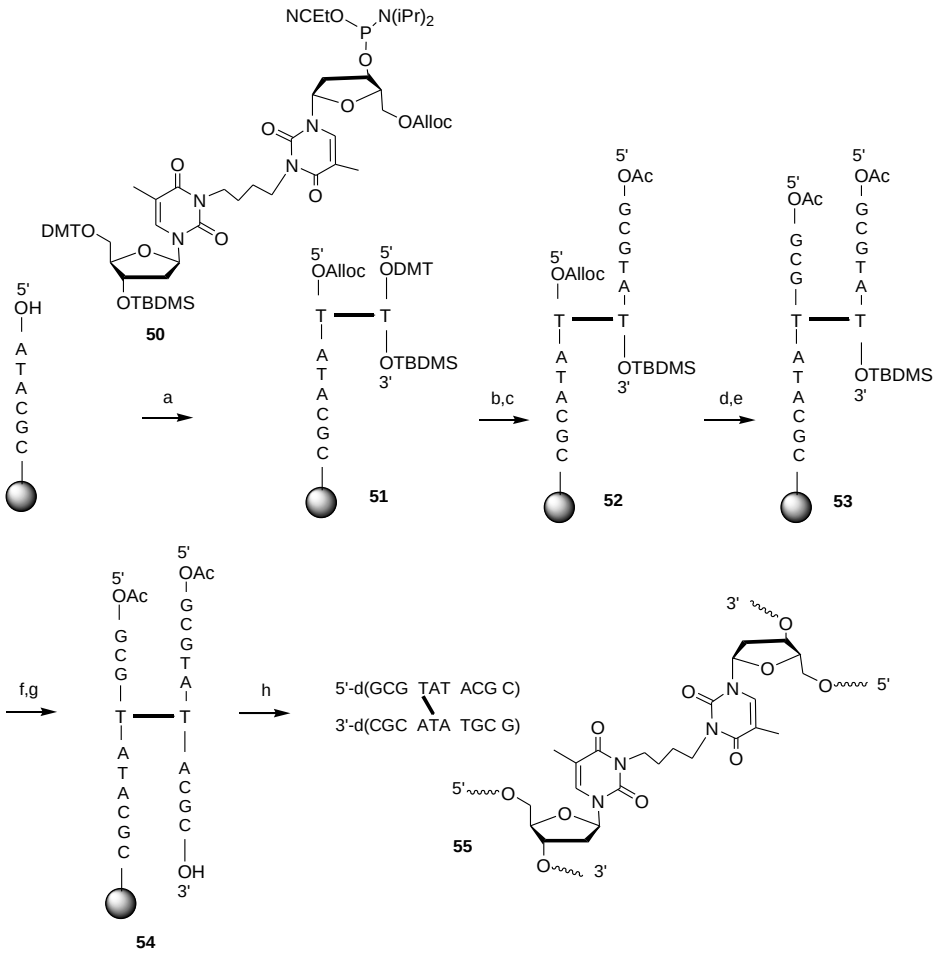
Key: (a) Coupling of **45** during solid phase oligonucleotide synthesis; (b) oligomer synthesis; (c) capping; (d) desilylation; (e) oligomer synthesis using 3'-cyanoethyl "reverse" phosphoramidites; (f) ammonolysis.

2.1.3.3. Synthesis of N3-thymidine-butylene-N3-thymidine ICL

In addition to a TBDMS protection on the 3'-hydroxyl, ICL dinucleotide phosphoramidite **50** was protected with an allyloxycarbonyl (Alloc) group on one of its 5'-termini (Scheme 26).[33] After incorporating the phosphoramidite using solid phase synthesis, **51** was converted to **52** upon continued synthesis and acyl end-capping. Once the Alloc was removed, solid phase synthesis using regular phosphoramidites and end capping yielded **53**. Finally, removal of TBDMS and synthesis using 3'-cyanoethyl "reverse" phosphoramidites led to **54,** that upon ammonolysis produced full length ICL product **55**.[33] Synthesis of N3-thymidine-butylene-N3-thymidine ICL (**55**) demonstrated that

using a greater number of orthogonal hydroxyl group protections provides more complete control over the DNA sequence.

Scheme 26. Synthesis of N3-thymidine-butylene-N3-thymidine ICL



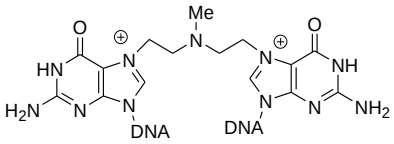
Key: a) Coupling of 50 during solid phase synthesis; b) DMT removal; c) solid phase synthesis with 5'-DMT 3'-cyanoethyl "regular" phosphoramidites and capping; d) ALLOC removal; e) solid phase synthesis with 5'-DMT 3'-cyanoethyl "regular" phosphoramidites and capping; f) TBDMS removal; g) solid phase synthesis with 3'-cyanoethyl "reverse" phosphoramidites; h) ammonolysis.

2.2. Structural effects of interstrand cross-link formation

ICL induced structural change has been investigated using various methods including interatomic distance calculated by NOESY,[88] proton exchange measured by HSQC,[89] and backbone deformation measured by ^{31}P NMR,[90] molecular simulation[1] and sometimes with the combination[91] of more than one of these methods. In other instances,

cross-linked DNA was treated with Maxam-Gilbert sequencing reagents to probe for backbone deformity.[40] Differences in UV thermal melting temperature between cross-linked and un-cross-linked duplexes serve as a measure of stability offered by cross-linking as well.[32] As shown below, a combination of these techniques has often been used to elucidate ICL structure.

2.2.1. Structural effect of mechlorethamine ICL (5) formation



N7G-N7G-EMA, **5**

EMA = ethylmethylamine

The structural effect of N7G-N7G-EMA (**5**) was first investigated by Hopkins employing molecular mechanics energy minimization computations, coupled with gel retardation to measure DNA bending.[1] The difference in cross-linking bridge length (7.5 Å) and the distance between the N7 of Gs in a 5'-(GNC) sequence (8.9 Å) forced the cross-linked DNA to depart from its canonical B form. Computation showed that the cross-link induced local structural distortions including buckling and propeller twist of trinucleotides encompassing the ICL. Especially, the cross-linking guanosines were reorganized. Additionally, reduced electrophoretic mobility of DNA containing multiple mechlorethamine ICLs within one or two turns (10.5 bp per turn) indicated 12° bending per lesion.[1] A consequence of this distortion was noticed when nitrogen mustard ICL mimics were synthesized by reductive amination of aldehydes (Scheme 12). Greater yields were obtained when tether length matched the G-G distance in 5'-(GNC) sequence.[29]

2.2.2. Structural effect of CENU ICL formation

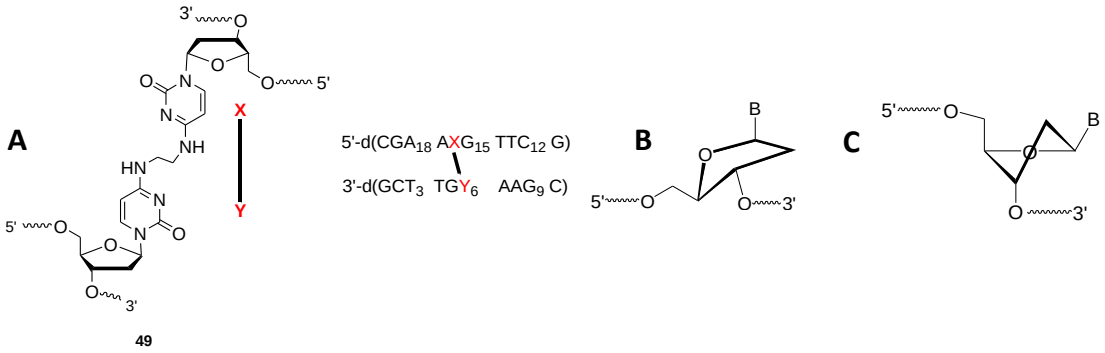


Figure 3. Structural probing N4C-ethyl-N4C ICL **(49)**. (A) Oligonucleotide used to determine the structural effect of **49**. (B) C4'-exo sugar pucker detected for cross-linked residues. (C) C2'-endo sugar pucker of B-form DNA.

N4C-ethyl-N4C ICL **49** in a 5'-(CG) sequence was used as a CENU mimic by the Miller lab to elucidate its structure (Figure 3).[91] ^{1}H NMR and restrained molecular dynamics showed that the ICL was accommodated within the major grove without affecting any base pairing interaction (Figure 3).[91] Enhanced propeller twisting was detected in cross-linked nucleotides, along with buckling between the bases within and around the ICL that gradually attenuated within three nucleotides from 5'-(CG) sequence (Figure 3).[91] Most notably, the conformation of the sugar bases of cross-linked nucleotide switched to C4'-*exo* pucker compared to the C2'-*endo* observed in B-form DNA (Figure 3B and C).[91] Formation of 5'-(CG) ICL stabilized the duplex. For 5'-(GAA C*GT TC) in PBS buffer (C* was the position of cross-linking), the T_m increased from 14.5°C to 81°C. Attempted structural decryption of **49** in the "unobserved" 5'-(GC) sequence context (by NMR) was unsuccessful due to considerable exchange around the cross-linking site, indicating instability.[91] This was further corroborated by thermal UV denaturation of duplexe with ICL **49** in 5'-(CG) sequence that showed significantly higher T_m compared to a comparable duplex containing 5'-(GC).[32]

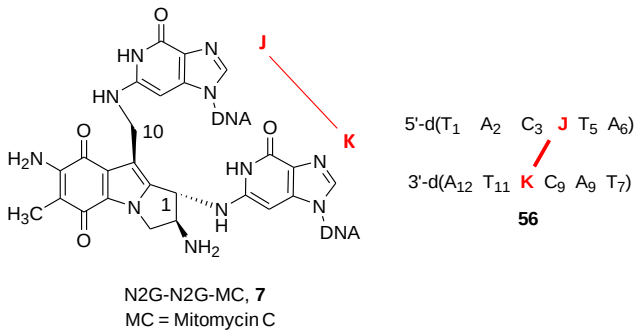


Figure 4. Sequence of Oligonucleotide (**56**) used to determine the structural effect of N2G-N2G-MC ICL (**7**).

2.2.2. Structural effect of mitomycin ICL formation

The solution structure of a mitomycin ICL (**7**, Figure 4) was divulged by computational studies aided with inputs from NMR interproton distance parameters obtained for oligonucleotide **56** (Figure 4).[88] Accommodating the mitomycin C (MC) ICL affected the two strands in a different manner but overall the cross-linked DNA did not deviate from B-form structure. Distance correlation identified the position of the adduct in a widened minor groove (9-10 Å compared to 3-7 Å for B-form DNA) and established the connectivity of N2G-MC(C10") and N2G-MC(C1") linkages.[90] The base pairing and stacking were unaffected (including the cross-linked C_3-G_4 step). Both G_{10}-T_{11} and G_4-T_5 sugar rings were in closer proximity than their counterparts in the noncross-linked duplex, with the former pair having a relatively smaller distance. In addition, differences in ^{31}P chemical shifts between duplex and cross-link revealed that the C_3-G_4-T_5 and C_9-G_{10}-T_{11} steps were perturbed differently. The aromatic ring of the mitomycin cross-link was aligned towards G_{10}-T_{11} instead of positioning at the center of minor groove. Furthermore, sugar puckers in G_{10} and T_{11} deviated from C2'-endo. The different perturbation of the two strands could possibly be attributed to an asymmetric chemical structure of the mitomycin ICL causing non-uniform stereoelectronic interactions with

DNA across the minor groove. The T_m of an un-cross-linked 10 bp duplex increased from 18°C to 50°C upon introduction of MC ICL in the middle of the sequence.[41] Furthermore, structural probing using chemical agents was carried out on cross-linked and un-cross-linked duplex using nucleotide specific sequencing reagents.[40] Hyperreactivity towards footprinting agents was highest at the cross-linked dGs and gradually diminished after three to four nucleotides.

2.2.3. Solution structure of ICL generated from an aldehyde

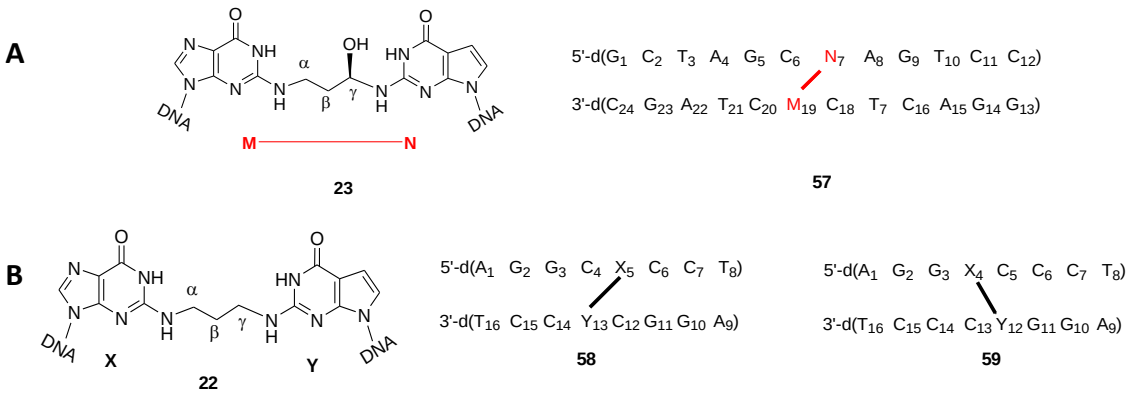


Figure 5. Determination of solution structure for ICLs generated by aldehydes. (A) Duplex **57** used for the structure determination of acrolein ICL (**23**). (B) Duplex **58** and **59** containing ICL **22** in 5'-(CG) and 5'-(GC) sequence, respectively.

Asymmetric deformities in the strands were observed when cross-linked duplex **57** was probed for distortion induced by acrolein ICL (**23**) in 5'-(CG) sequence using NMR (Figure 5A).[92] The chemical shift change of deoxyribose C1' and imino protons of nucleobases were compared between **57** and un-cross-linked duplex. The exchangeability of the base imino protons (probed by change in peak broadness with temperature) was investigated for the two substrates to study the stability of base pairs. In addition, the distances between aromatic nucleobase protons and C1' protons for the two substrates (detected by NOESY) was compared to detect cross-link induced changes in the

backbone. The minor groove ICL only minimally affected the H-bonding for the cross-linked nucleotides, with no observed effect on base pairing beyond the C_6-Y_{19} and X_7-C_{18} pairs. In the top strand, backbone deformities were highest for C6 and X7, dissipated towards G5 and G9 and then completely disappeared. Distortion was detected in nucleotides T_{17}, C_{18}, Y_{19} and C_{20} in the bottom strand, with the latter three being most deformed.

Trimethylene ICL **22**, the reduced analogue of the acrolein ICL (**23**), was inserted in the native 5'-(CG) (**58**) and 5'-(GC) (**59**) sequences to probe the selective formation of the ICL in the former (Figure 5B).[93] In the former sequence, the cross-linked X_5 N2 was in gauche conformation with Cγ, while in the latter they were in an unfavorable half gauche conformation. In addition, the X_4-C_{13} and C_5-Y_{12} pairs completely lost π stacking interactions with neighboring bases by roll and twist in ICL **59**, destabilizing the structure. This was supported by UV thermal melting of a trimethylene ICL in a 5'-(GC) sequence and in a 5'-(CG) sequence, where the latter showed a significantly higher T_m.[74,94]

2.2.4. Structural effect of psoralen ICL formation

Solution phase NMR was employed to explore the structure of 4'-(hydroxymethyl)-4,5',8-trimethylpsoralen (HMT) induced cross-linked duplex **60** (Figure 6).[89] The T_4 in the top strand was involved in formation of the furan-side adduct while T_{12} in the bottom strand was part of the pyrone-side cross-linking. Like aldehyde and mitomycin ICLs, HMT adduct formation affected the two strands differently, with the distortion being more on the pyrone-side. The H-bonding in the base pairs encompassing the cross-link were minimally affected. The NH imino protons in T_{12} and G_{11} were more

prone to temperature induced broadness than T_4 and G_3 bases. The distance between the cross-linked Ts with the purine moieties in their 5'-adjacent dGs was reduced after cross-linking, enhancing π-stacking. No bending in the backbone was observed as a result of ICL formation. In both strands the base pairing returned to normal within three nucleotides from the site of adduct formation. Overall, the cross-linked structure was significantly more stabilized than the un-cross-linked duplex due to enhanced base stacking and intercalation of the psoralen moiety. As a result, introduction of a psoralen cross-link in the middle of the strand increased of T_m of a 16-mer duplex from 51 °C to more than 75 °C.[95]

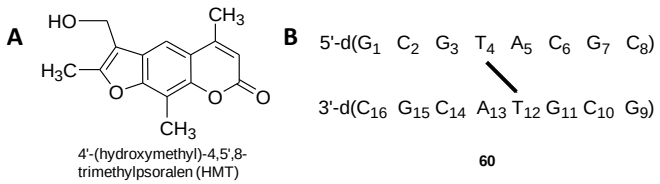


Figure 6. Structures of HMT molecule and cross-linked duplex used for investigating psoralen cross-link structure. (A) Chemical structure of HMT molecule. (B) Sequence of HMT cross-linked duplex, with the cross-link formed between T_4 and T_{12}.

2.3. Repair of interstrand cross-links using UvrABC

The SOS response is actuated when a bacteria is exposed to stress. In this state, many defense and repair proteins are produced from genes, which otherwise remain in a dormant state.[96] SOS induced protein expression and related nucleotide excision repair was first observed in 1964, when excision of damaged DNA strands containing a thymine dimer[97] was detected in UV irradiated cells.[98] The excision was mediated by three genes controlled through the SOS response, *uvrA*, *uvrB* and *uvrC*.[98] Articles published during the mid-seventies reported repair of bulky lesions by extracted gene products from the above mentioned genes in *E. coli*.[99,100] In 1983, Aziz Sancar isolated the enzymes and

coined the term UvrA, UvrB, and UvrC for these repair proteins in *E. coli*.[101] The UvrABC repair system recognizes and repairs a diverse collection of lesions, indicating that chemical structure of the damage probably plays minimal role in recognition. Rather, backbone deformity induced by a lesion has been revealed to play a significant role in damage recognition, and incision efficiency.[102] Presence of ICLs affects both strands and requires the repair proteins to make even more complex choices than for lesions on a single strand, making cross-link forming agents of the most potent genotoxins according to some researchers.[103] Herein we discuss the effect of damage induced distortion on recognition and incision in the context of all three proteins.

2.3.1. Role of UvrA

UvrA, the 105 kDa ATPase enzyme acts as a dimer ($UvrA_2$) during primary damage recognition.[104,105] A recently published crystal structure shed light on the binding and damage recognition of a modified DNA [two fluoresceinylated Ts (F) present in palindromic 5'-FCCGGA sequence] by UvrA (Figure 7A).[106] The dual modifications resulted in a 15° bend and 20° unwinding in the sequence encompassing the fluoresceinylated T (Fl-T) bases along with greater shear (movement of H-bonded bases in the plane) and stretch (span between the bases along helical axis).[106] A previously identified concave semicylindrical lane on the surface of the UvrA dimer ($UvrA_2$) accommodated the DNA in a non-specific manner (Figure 7B and 7C).[107] Two zinc finger motifs (one from each UvrA monomer) were seen approaching the DNA backbone (Figure 7D). The unwound and bent region placed in the middle of the lane and probably serves as the primary damage reporter to UvrA.[106] It was speculated that the clash of the DNA binding region with the bulged region of a damaged DNA acted as the signal for

damage, whereas this steric conflict would be absent for undamaged DNA. Deleting the zinc fingers affected the differentiation capability of UvrA and they were thus speculated

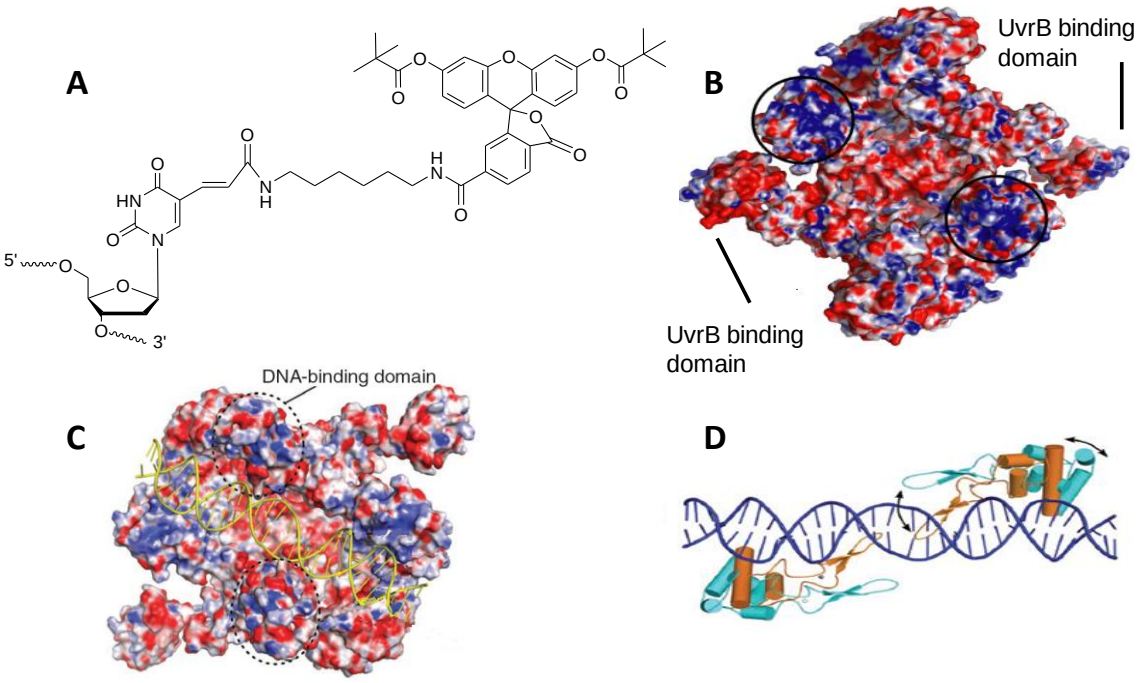


Figure 7. DNA damage recognition by UvrA as visualized in crystal structure.[106,107] (A) Structure of fluoresceinylated T (Fl-T). (B) Surface representation of the UvrA dimer ($UvrA_2$) with surface potential in red (negative) and blue (positive) (PDB 2R6F). Region containing charged amino acids (circled) and UvrB binding domain are indicated.[107] (C) $UvrA_2$ in complex with DNA (yellow) (PDB 3PIH).[106] (D) Non-specific interaction of zinc finger motifs with damaged DNA (blue) (PDB 3PIH). Zinc atoms are shown in grey.[106] Images are reproduced without permission from the work of Jaciuk[106] and Pakotipraha.[107]

to play some role in damage recognition.[106] Based on the mutation studies both hydrophobic and charged residues present at the end of the lane were deemed responsible for damage recognition (Figure 8A). The length (100 Å) of the lane was in accordance with UvrA footprinting (33 bp).[108] The binding of UvrA to DNA was accompanied by partial unwinding of the duplex, although the exact reason or significance of this is unknown.[109]

2.3.2. Role of UvrB, damage recognition and dynamics of preincision complex formation

After the initial scanning, the complex between UvrA dimer and DNA ($UvrA_2$-DNA) further associates with UvrB dimer ($UvrB_2$) and forms the $UvrA_2UvrB_2$-DNA heterotetramer protein-DNA complex for further damage localization. A set of recently published crystal structures elucidated how the complex distinguished between damaged and undamaged DNA (Figure 8).[110] In the $UvrA_2$-DNA complex, the protein remained in an "open tray" like conformation. In this conformation, the damaged DNA does not closely interact with the surface of shallow and wide (43 Å) semicylindrical lane of UvrA (Figure 7C and 8A). As a result of the weak interaction, UvrA alone has a low specificity for damaged DNA (2-5 fold over native DNA)[111] and comparable affinity for a wide range of lesions (K_d = 1-14 nM).[112,113] However, the lane becomes narrow (30 Å) and deeper in the "closed groove" heterotetramer structure (width 30Å, Figure 8B).[110] This conformation change ("open tray" to "closed groove") can be attained uninhibitedly for a native DNA (Figure 10C).[110] The $UvrA_2B_2$ complex then reversibly disassembles from the native DNA and $UvrA_2$ returns to the "open tray" structure (Figure 8C). The $UvrA_2$ then resumes its search for damaged DNA. However, the backbone deformation of a DNA lesion impedes this conformational change ("open tray" to "closed groove"), generating a steric clash, and indicating the presence of damage (Figure 8D).[110] The subsequent events leading to UvrC binding are still debated. It was proposed that the UvrB molecules translocate from the 5' to 3' direction, displacing the A_2 dimer and converge on the damage site to form the preincision complex (Figure 8E).[115] However, it

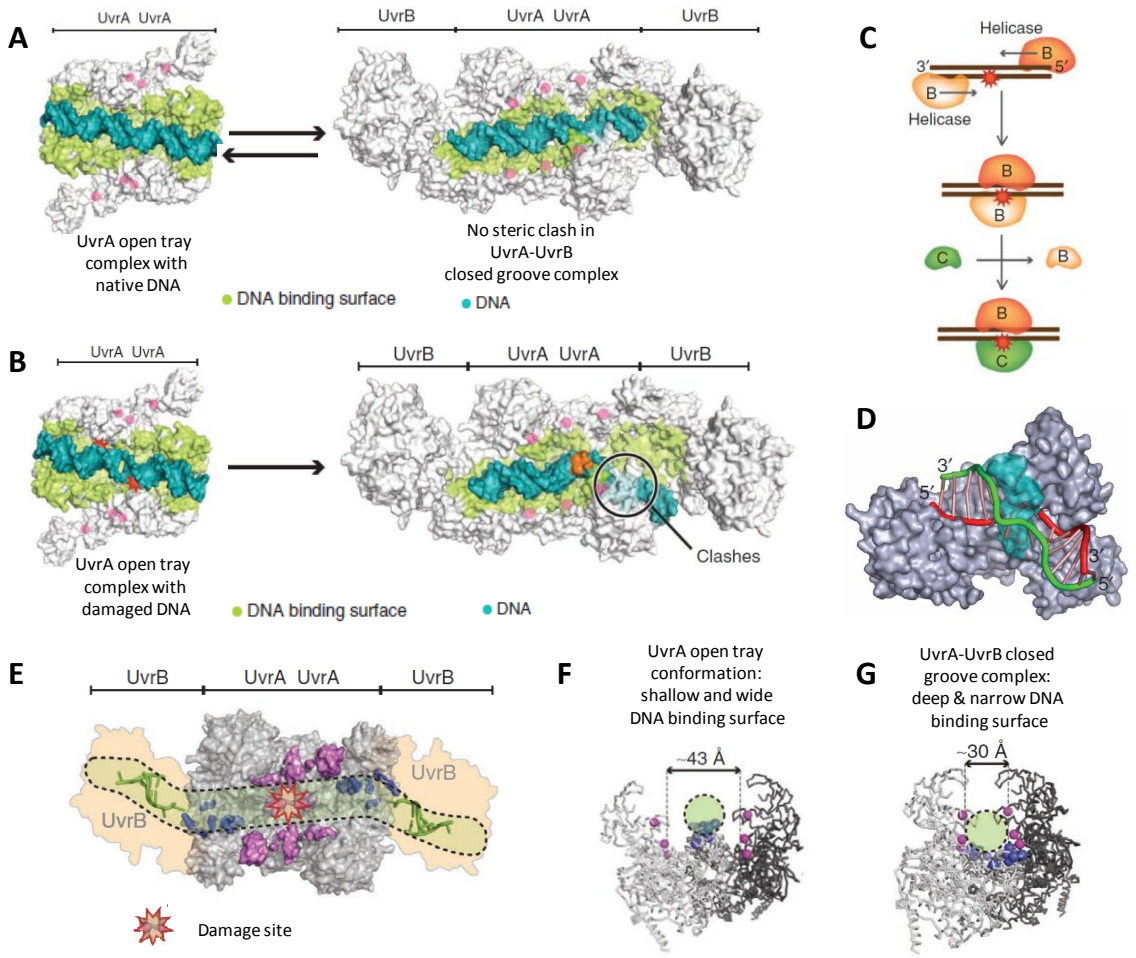


Figure 8. Damage detection by A_2B_2 heterotetramer and formation of preincision complex. (A) Location and width of DNA binding surface in $UvrA_2$ "open tray" conformation.[110] (B) Location and width of DNA binding surface in $UvrA_2B_2$ closed groove conformation.[110] (C) Model of uninhibited transition from A_2 "open tray" to A_2B_2 "closed groove" conformation for a native DNA. DNA binding surface and native DNA are shown in green and cyan, respectively.[110] (D) Model of transition from A_2 "open tray" to A_2B_2 "closed groove" conformation for a damaged DNA. DNA binding surface and native DNA are shown in green and cyan, respectively.[110] (E) Proposed model for UvrB localization on damage and preincision complex formation.[114] Lesion containing strand, undamaged strand, and β hairpin are shown in green, red, and cyan. (G) Proposed path of damaged DNA on A_2B_2 complex.[110] Images are reproduced without permission from the work of Pakotipraha.[110]

is unclear whether the preincision complex was comprised of one or two molecules of UvrB, and evidences for both models have been reported.[116,117] In the complex, the nucleic acid is wrapped around the UvrB surface, further aiding local DNA unwinding.[118] In a model of the preincision complex based on a crystal structure of a hairpin DNA and a lone UvrB molecule, insertion of a β hairpin from the latter into the helix caused local

unwinding and base flipping (Figure 8F).[114] The damaged base remained on the outer DNA strand while extrahelical bases on the inner strand were flipped and inserted into a hydrophobic pocket behind the β hairpin in UvrB.[114,119] Transfer of DNA from the "open tray" $UvrA_2$-DNA to "closed groove" UvrB-UvrA-UvrA-UvrB heterotetramer (groove length length 145 Å, Figure 10G) to UvrB (68 kDa, Figure 10F) is consistent with changes in footprinting from 30 bp to 45 bp to 19 bp.[108,111]

2.3.3. Role of UvrC and strand selection for ICLs

After the presence of damage is confirmed by UvrA and UvrB, repair advances by recruitment of UvrC, which incises the lesion containing strand 5' and 3' to the damage. The incisions take place at the fourth or fifth phosphodiester bond 3' to the damage (third phosphodiester bond for ICL[120]), and eighth phosphodiester bond 5' to the damage (ninth phosphodiester bond for ICL[120]) both generating 3'-OH and 5'-phosphate (Figure 9).[121] While a single strand nick is recognized as a substrate by UvrABC, its incision is not properly characterized.[122,123] In a substrate containing a single strand nick, UvrABC incision occurs at sixth phosphodiester bond 5' to the nick.[123] However, no study addressed the UvrABC incision 3' to a nick. The N-terminal and C-terminal sides of UvrC carry out the 3'- and 5'-incisions respectively.[121] For monoadducts, the incision is followed by removal of the incised oligonucleotide patch by UvrD helicase activity ("unhooking"), subsequent repair synthesis and ligation (by DNA polymerase and ligase).[121] For interstrand cross-links, the first strand repaired is followed by another round of incision of the remnants of the ICL on the other strand, "unhooking" by UvrD, and repair synthesis.[18]

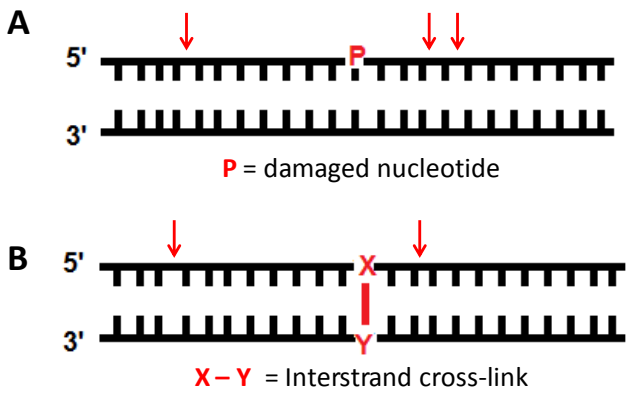


Figure 9. UvrC incision locations on cross-linked and un-cross-linked substrates. (A) UvrC cleavage sites for the incision of oligonucleotide containing a damaged nucleotide (**P**).[121] (B) UvrC cleavage sites for the incision of oligonucleotide containing an interstrand cross-link (X – Y).[121]

2.3.4. Effect of distortion and adjacent sequence on preincision complex dynamics and incision yield

Gel retardation assays have been extensively used to analyze protein-DNA interactions in the UvrABC pathway. DNA binding affinities of UvrA, UvrAB and UvrB have been juxtaposed against the size of the lesion, helical distortion, and incision efficiency to investigate their interrelation. Due to its weak interaction with damaged DNA, UvrA affinity for assorted lesions that induce a range of distortion often does not parallel incision yields. For instance, *o*-benzylhydroxylamine-modified apyrimidinic (BA) sites were incised more efficiently than abasic sites due to the larger size of the former despite UvrA binding being similar (K_d = 14 nM and 6 nM, respectively) (Figure 10A and B).[124] In contrast, recognition of DNA by UvrAB or UvrB in the form of a preincision complex can be translated more reliably into incision yield. The higher affinity of UvrAB for BA lesion (K_d = 80 – 120 nM and 4 nM for AP and BA lesions, respectively) paralleled the greater incision yield for BA over AP lesions. Therefore, the preincision complex formation was proposed as the rate determining step of bacterial NER.[124] Similar correlation between UvrAB affinity, incision efficiency, and lesion size

was established for bulky spiroiminodihydantoin adducts and the smaller guanidinohydantoin lesion as well (Figure 10C and D).[125] However, as shown for similar-

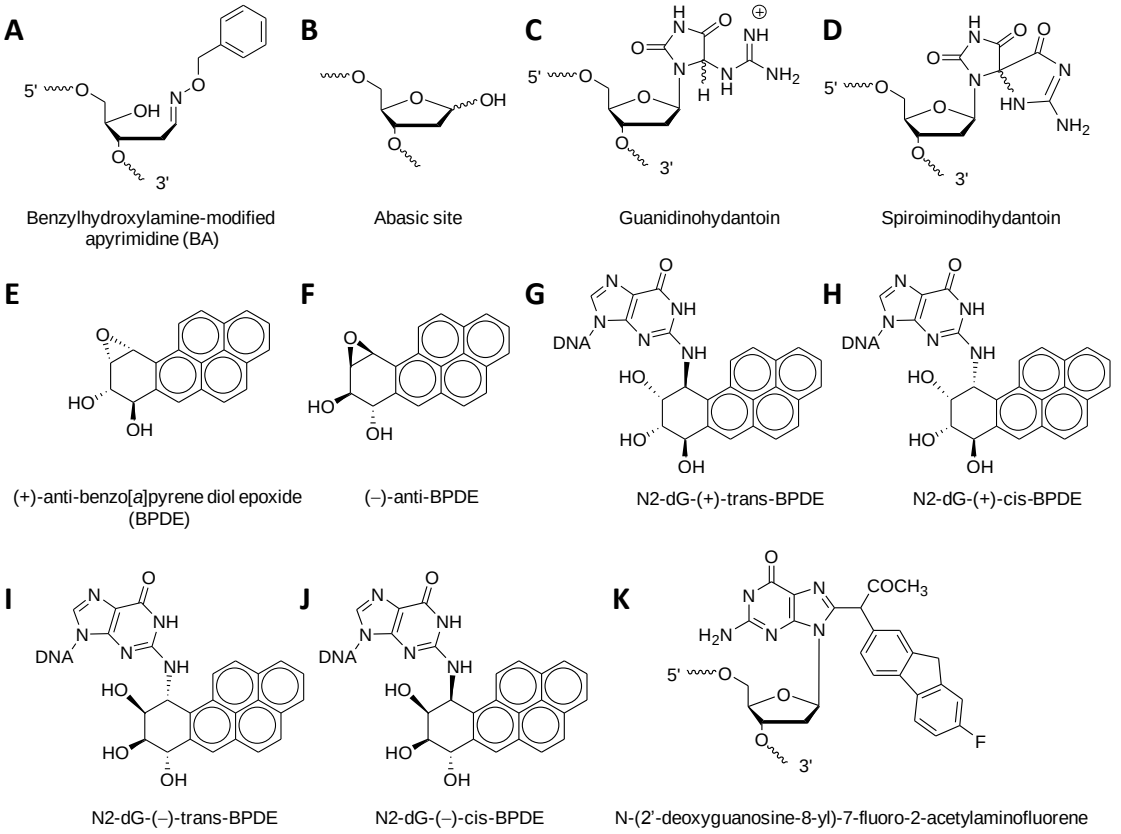


Figure 10. Chemical structures of various electrophiles and their nucleobase adducts. (A) Benzylhydroxylamine-modified apyrimidine (BA). (B) AP site. (C) Guanidinohydantoin. (D) Spiroiminodihydantoin. (E) (+)-Anti-benzo[*a*]pyrene diol epoxide (BPDE). (F) (–)-Anti-BPDE. (G) N2-dG-(+)-trans-BPDE. (H) N2-dG-(+)-cis-BPDE. (I) N2-dG-(–)-trans-BPDE. (J) N2-dG-(–)-cis-BPDE. (K) N-(2'-deoxyguanosine-8-yl)-7-fluoro-2-acetylamino-fluorene.

sized anti-benzopyrene diol epoxide (BPDE) adducts of G, interrelation of preincision complex, incision yield, and inherent backbone deformity (unrelated to size) could be less straightforward (Figure 10E – J). BPDE compounds reacted with G to generate four different adducts, N2-G-(+)trans-, (–)trans-, (+)cis-, (–)cis-BPDE.[126] The cis adducts induced greater backbone distortion by intercalation of the aromatic core.[127] In contrast, due to placement of the adduct in the minor groove, the trans adducts were less deformed with all the base pairing remaining intact.[127] While the cis adducts were more efficiently

incised, no clear distinction was visible in terms of preincision complex formation.[126] Larger lesion size and higher backbone deformity may be linked to greater thermodynamic destabilization, since the latter would ease local unwinding seen in the preincision complex. The effect of thermodynamic destabilization was emphasized during repair of N-(2'-deoxyguanosine-8-yl)-7-fluoro-2-acetylaminofluorene adducts at G_1, G_2, or G_3 in 5'-d(CC$G_1G_2CG_3$CC) (Figure 10K).[102] Among the three, greater destabilization (characterized by thermal melting) for the G_1 and G_3 adducts were consistent with their more efficient incision compared to G_2.

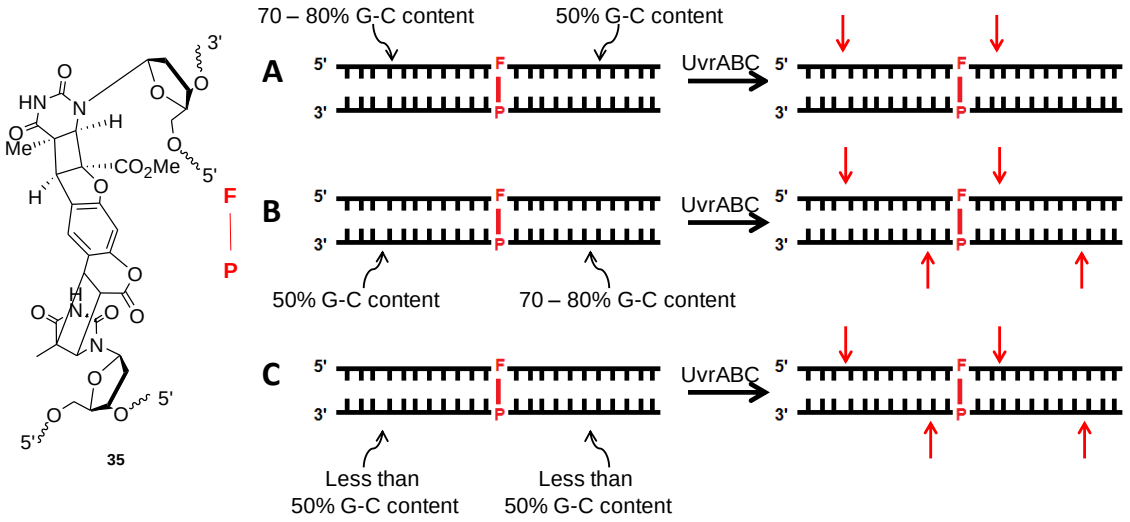


Figure 11. Influence of G-C content of an oligonucleotide containing a psoralen ICL (**35**) on the strand specificity of UvrABC incision. (A) Incision profile when 5'-side and 3'-side of the furan adducted strand (F) has 70-80% and 50% G-C content, respectively. (B) Incision profile when 5'-side and 3'-side of the pyrone adducted strand (P) has 70-80% and 50% G-C content, respectively. (C) Incision profile when 5'-side and 3'-side of either strand has less than 50% G-C content.

A related unresolved issue is the strand preference during incision of asymmetric ICLs (different chemical structures at the termini). Although the strand preference for psoralen ICL has been linked with distortion and adjacent sequence, no clear trend has been found. The initial findings of exclusive furan side bimodal incision could be correlated with greater distortion seen at the furan side of ICL.[89,120] Further research

demonstrated that the presence of more than 70-85% G-C content on 5' side and 50% G-C content on 3' side of the furan adducted strand led to predominant or exclusive incision

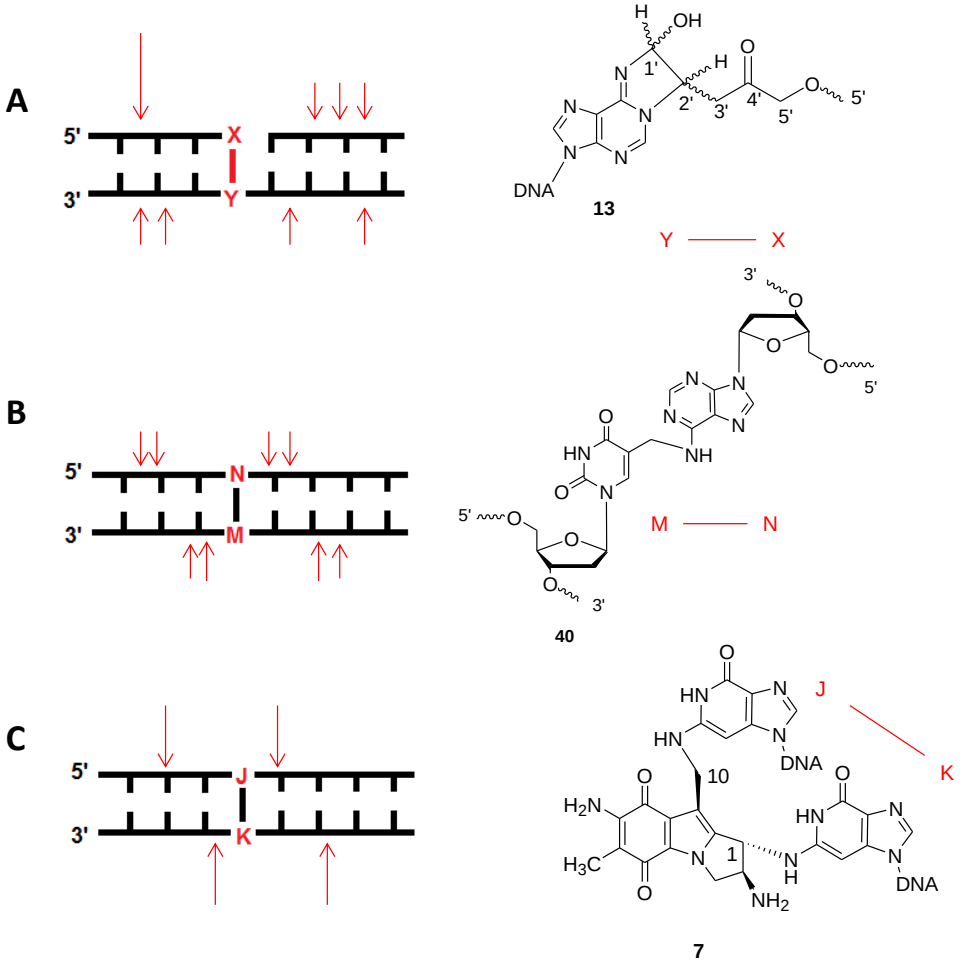


Figure 12. Generation of double strand break by UvrABC incision on C4-AP thermodynamic ICL (**13**), A-T ICL (**40**) and mitomycin C ICL (**7**). (A) Cartoon showing UvrABC incision of a substrate containing **13**.[22] (B) Cartoon showing UvrABC incision of a substrate containing **40**.[21] (C) Cartoon showing UvrABC incision of a substrate containing **7**.[20]

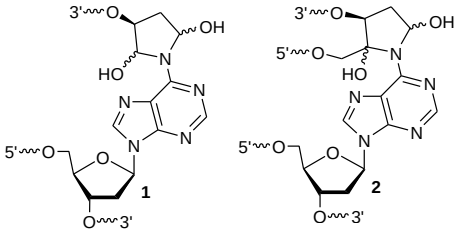
on that side (Figure 11).[128] However, similar base composition on the pyrone adducted strand, or less than 50% G-C content on either side caused incisions on both strands without any clear preference.[128] Unfortunately, neither of the two independent studies on the repair of the mitomycin C cross-link (another asymmetric ICL) have addressed the issue of strand specificity.[19,20]

Importantly, analysis of UvrABC treatment of C4-AP thermodynamic ICL (**13**), A-T ICL (**36**) and mitomycin C ICL (**7**) detected double strand break (DSB) generation. C4-AP thermodynamic ICL **13** is accompanied by an adjacent strand break and is therefore classified as a clustered lesion (more than one damage site within 1 – 1.5 helical turns).[31] The DSB detected during UvrABC treatment of **13** occurred as a result of incision on the strand containing the cross-linked A (Figure 12A).[22] In contrast, the DSBs formed during UvrABC incision of **36** and **7** results from repetitive cleavages on both strands (Figure 12B and C).[20,21]

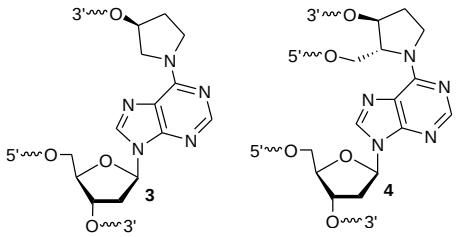
3. Results and discussion

3.1. Synthesis of ICL mimics

3.1.1. Synthesis of cross-linked oligonucleotides containing DOB ICL mimic (3)



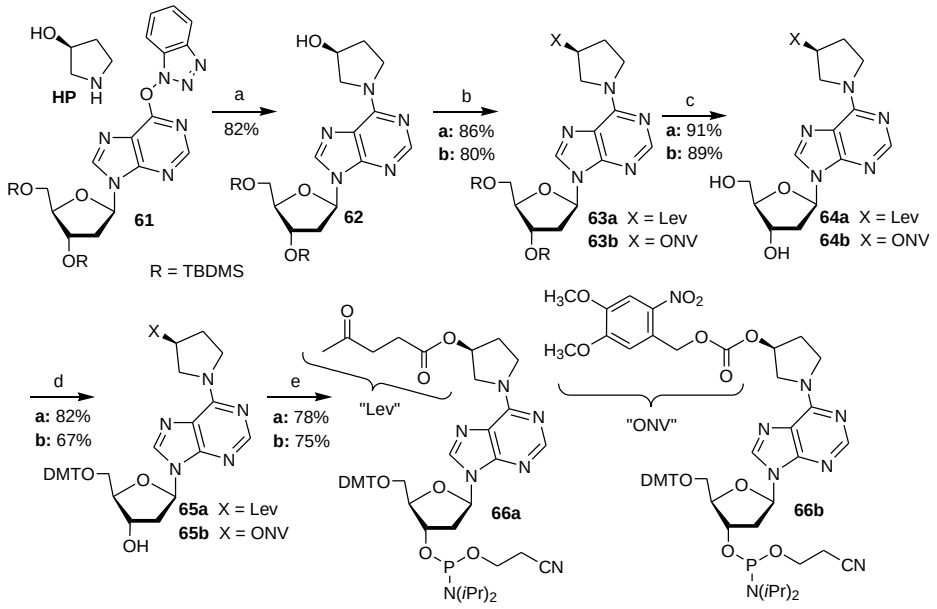
In order to stabilize the DOB and C4-AP kinetic cross-links (**1** and **2**), analogues of the reduced versions of the native ICLs were synthesized, where the unstable hemiaminal functionalities were removed (**3** and **4**). An orthogonal protecting group strategy, in combination with solid phase oligonucleotide synthesis approach was utilized to synthesize the ICL mimics. The strategy necessitated one protecting group for the DOB and two for the C4-AP analogue, that would allow further extension from their respective pyrrolidinyl hydroxyls during solid phase synthesis. The protecting groups needed to be robust enough to survive the standard solid phase DNA synthesis conditions but cleavable without any loss of various protecting groups on commercial phosphoramidites. We experimented with a number of commercial DNA phosphoramidite protectors and hydroxyl masking groups, as well as various demasking conditions to optimize strategies for making the DOB and C4-AP ICL analogues.



3.1.1.1. Synthesis of Lev-DOB and NVOC-DOB ICL analogue phosphoramidites

Synthesis of the oligonucleotide containing **3** involved making the "template" strand carrying the cross-linked A portion of **3** using commercially available 5'-dimethoxytrityl-protected 3'-cyanoethyl (CE) "regular" phosphoramidites. The requisite phosphoramidites (**66a**,**b**) were synthesized from an activated form of 2'-deoxyinosine (**61,** Scheme 27). The hydroxypyrrolidine (HP) substitution was carried out by employing the S_NAr chemistry developed by Lakshman.[129] The substitution product was utilized to install the levulinyl and nitroveratryloxycarbonyl protecting groups to afford **63a** and **63b** respectively. After removing the silyl groups, the nucleoside analogues (**64a**,**b**) were dimethoxytritylated (**65a**,**b**) and phosphitylated via standard procedures to produce the phosphoramidites.

Scheme 27. Synthesis of Lev-DOB and NVOC-DOB ICL mimic phosphoramidites

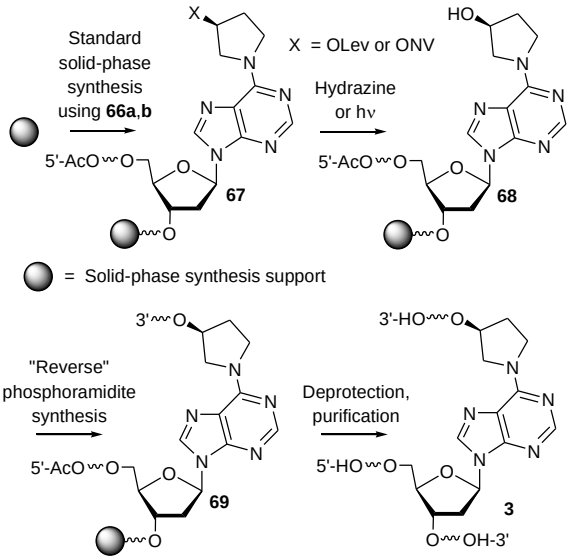


[a]Key: a) HP, Cs_2CO_3, THF b) Levulinic acid, EDCI, DMAP, DMF or *o*-nitroveratryl chloroformate, DMAP, CH_2Cl_2
c) $Et_3N{\bullet}3HF$, THF d) DMTCl, pyridine e) 2-Cyanoethyl chlorophosphoramidite, diisopropylethylamine, CH_2Cl_2

3.1.1.2. Incorporation of Lev-DOB phosphoramidite

3.1.1.2.1. Stability of acetyl and pivaloyl capping and standard phosphoramidite amine protecting groups under delevulinylation condition

Scheme 28. Synthesis of DOB ICL mimic 3 from phosphoramidites 66a,b



The compatibility of various standard protecting groups used in phosphoramidite based solid phase DNA synthesis needed to be verified for the synthesis of the DOB ICL mimic **3** using an orthogonal deprotection strategy. During the incorporation of **66a** into support bound oligonucleotide (**67**), the best coupling yields (50-98%) were obtained when **66a** was double-coupled (concentration 0.15 M) for 15 min (Scheme 28). Following removal of the levulinyl group to free the pyrrolidinyl hydroxyl (**68**), 5'-cyanoethyl "reverse" phosphoramidites were employed to prepare the opposing strand or "arm" (**69**). Cleavage of **69** from the support by ammonolysis produced DOB ICL analogue **3**. To avoid unwanted branching during the synthesis of the opposing or "non-template" strand, complete 5'-OH protection (capping) and its stability under delevulinylation conditions was essential. Following detritylation, the 5'-OH of the

template strand was manually acetyl capped using a solution containing 5% each of acetic anhydride (vol/vol), N-methylimidazole (weight/vol), and pyridine (vol/vol) in THF following detritylation. The resin was subjected to typical capping steps including mixing of reagents (5 s), passing the capping solution through the support (10 s), and waiting (10 s). These steps were repeated for 10 consecutive cycles to ensure complete capping.

We first attempted to synthesize cross-linked poly T oligonucleotide (**72**) containing **3** from resin bound **70** (Figure 13). Three different delevulinylation procedures, 0.5 M hydrazine in 4:1 pyridine/acetic acid (room temperature, 10 min), 0.25 M hydrazine in 3:2 pyridine acetic acid for 10 and 20 min (room temperature) were tested. Following delevulinylation, six nucleotides added via solid phase synthesis comprised the opposing strand ("arm"). The 5'-cyanoethyl "reverse" phosphoramidites were coupled for 15 min each to ensure complete conversion of **68** to **69** (Scheme 28). The synthesis was monitored by deprotecting the oligonucleotide after each step, loading them on preparative denaturing 20% PAGE, and visualizing them using a UV lamp

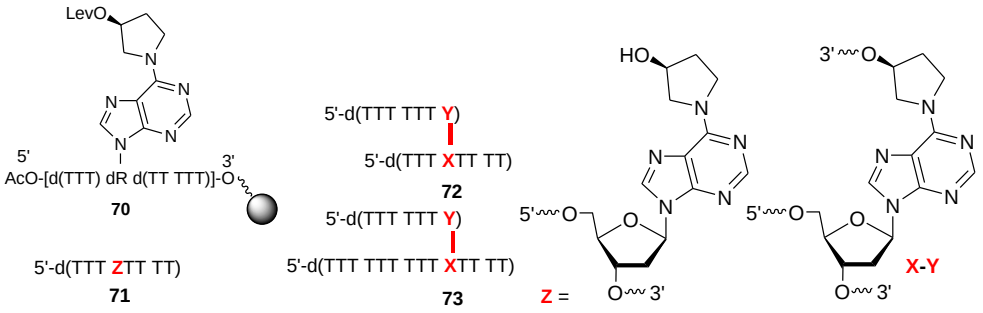


Figure 13. Sequence of the resin bound oligonucleotide **70**, template strand **71**, desired DOB ICL **72** and proposed sequence of the oligonucleotide with 12 T extension (**73**).

and fluorescent TLC plate (Figure 14). In addition to the desired **72**, a 12 nucleotide extension product was detected using any of the three delevulinylation conditions (Figure 14). The starting material **71** (from unmodified **70**) was not observed. Formation of **73**

was highest for the delevulinylation treatment using 0.5 M hydrazine in 4:1 pyridine/acetic acid (10 min, room temperature) and lowest for 0.25 M hydrazine in 3:2 pyridine/acetic acid (10 min, room temperature) (Figures 14B and C). We suspect that inefficient capping could be responsible for the generation of undesired extension, which was aggravated further during the treatment with higher hydrazine concentration (0.5 M). Characterization of the gel purified template strand (**71**),

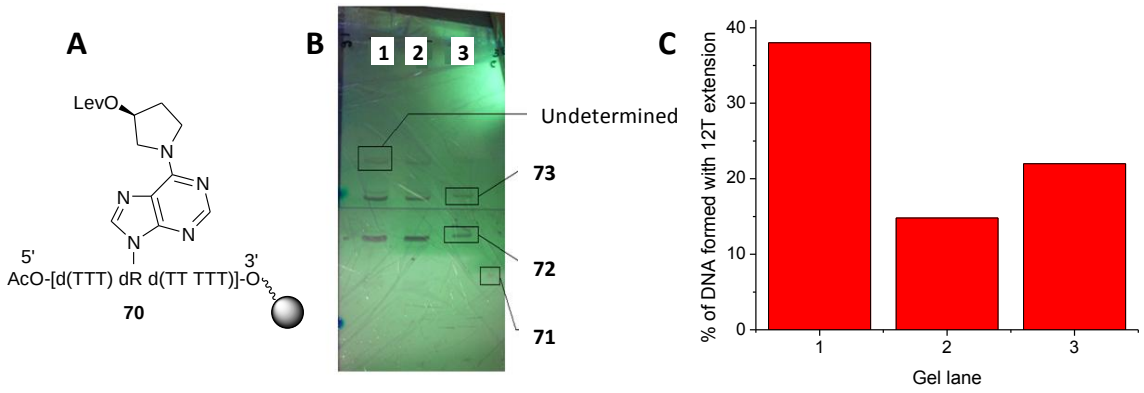


Figure 14. Stability of acetyl capping under hydrazine treatment. (A) Resin bound *O*-acetyl capped oligonucleotide **70** with levulinyl protection on pyrrolidine hydroxyl. (B) Preparative PAGE gel visualized with UV lamp and fluorescent TLC plate showing treatment of **70** with various hydrazine treatment conditions followed by 6T extensions. Positions of **71** (from unmodified **70**), **72** and **73** are annotated. Lane 1, treatment of **70** with 0.5 M hydrazine in 4:1 pyridine/acetic acid for 10 min. Lane 2, treatment of **70** with 0.25 M hydrazine in 3:2 pyridine/acetic acid for 10 min. Lane 3, treatment of **70** with 0.25 M hydrazine in 3:2 pyridine/acetic acid for 20 min. (C) Percentage of **72** isolated with respect to total extension product (**72** and **73**) formation.

cross-link (**72**) and unwanted extension product (**73**) were done by MALDI-TOF mass spectrometry (App. Figure 97 – 99). Position of **3** in **72** at the fifth nucleotide from 3'-terminus was confirmed by T sequencing of 3'-^{32}P-**72** (Figure 15).

To improve the 5'-hydroxyl protection, acetyl and pivaloyl protections were applied in modified manual capping methodology where the resin was incubated 10 times with the capping reagents for 4 min (as opposed to 10 s during the previous capping method) each cycle. The capped resin was treated with 0.25 M hydrazine in 3:2 pyridine/acetic acid (10 min, room temperature) followed by solid phase extension.

Oligonucleotides were subsequently deprotected and cleaved from the support by ammonolysis. An aliquot from the crude mixture was 5'-^{32}P-labeled, and analyzed by

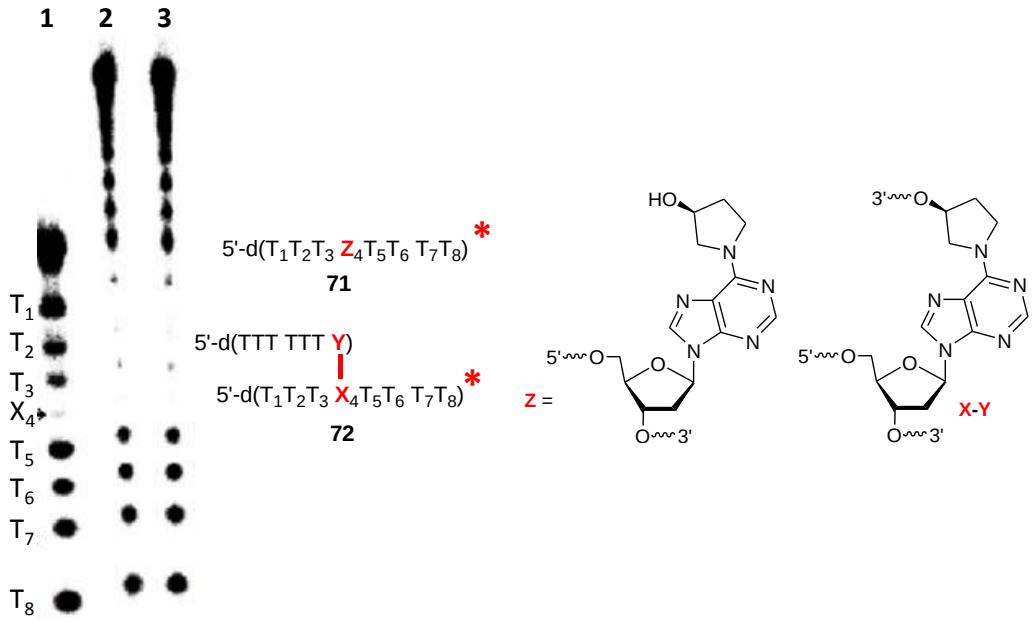


Figure 15. ICL location determination by T sequencing using $KMnO_4$. Lane 1, T sequencing of 3'-^{32}P-**71**. Lane 2 and 3, T sequencing of 3'-^{32}P-**72** for 8 and 15 min, respectively. Positions of the nucleotides are annotated.

20% denaturing PAGE. (App. Figure 1). The pivaloyl protection was partially removed after delevulinylation while the acetyl was unaffected. The efficiency of acetyl capping and delevulinylation using 0.25 M hydrazine in 3:2 pyridine/acetic acid (10 min, room temperature) was reaffirmed by resynthesizing **72**. After delevulinylation and synthesis of the arm using 5'-dimethoxytrityl 3'-cyanoethyl "regular" phosphoramidites, the oligonucleotides were deprotected and cleaved from support by ammonolysis. Analysis of a ^{32}P-labeled aliquot from the crude by 20% denaturing PAGE showed complete absence of **73** or **71** (App. Figure 2) and 94.3% conversion into the **72** (yield calculated based on the band intensity of **71** and **72**).

After the 5'-OH protection (capping) and delevulinylation steps were optimized, we sought to test the stability of standard phosphoramidite exocyclic amine protections

under hydrazinolysis conditions. Commercially available "regular" 5'-dimethoxytrityl (DMT) 3'-*β*-cyanoethyl (CE) phosphoramidites containing standard amine protections (benzoyl-A, isobutyryl-G, and benzoyl-C) and **66a** were employed to prepare the template strand (the strand carrying the cross-linked A) of **3**. N-benzoyl-dC was used instead of N-acetyl-C to avert deprotection during hydrazine treatment (Jonathan T. Sczepanski private communication).

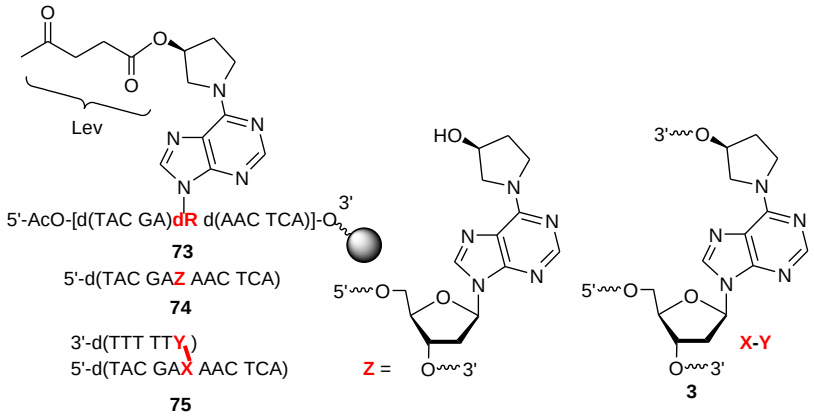


Figure 16. Sequences of resin bound capped oligonucleotide **73**, template strand **74** and corresponding DOB ICL mimic **75**.

Synthesis of ICL **75** was attempted to establish the feasibility of this approach (Figure 16). Following template strand synthesis and capping of the resin-bound oligonucleotide (**74**), three different delevulinylation procedures (0.25 M hydrazine in 3:2 pyridine/acetic acid for 1, 5, and 10 min, room temperature, respectively) were tested. As described before, the oligonucleotides were analyzed following extension using 5'-CE "reverse" phosphoramidites, deprotection, and cleavage from the resin. After loading on preparative denaturing 20% PAGE, the extension products were visualized using a UV lamp and fluorescent TLC plate (App. Figure 3). While none of these treatments caused any unwanted extension (branching), complete consumption of starting material was obtained only for the 10 min treatment. We therefore decided to adopt 0.25 M hydrazine

in 3:2 pyridine/acetic acid (10 min, room temperature) as the standard delevulinylation treatment. Gel purified template strand (**74**) and the cross-link (**75**) were characterized by MALDI-TOF MS (App. Figure 100 and 101).

3.1.1.2.2. Modification of capping, delevulinylation and coupling durations for template strand length ≥ 25 nucleotides

After establishing a standard capping and delevulinylation strategy, we attempted the synthesis of longer cross-linked oligonucleotides **77** and **79** containing DOB ICL analogue **3** (Figure 17). Due to greater length of the final products, the syntheses were done on 1000 Å resin. To maximize the yield, native phosphoramidites were double-coupled during the automated solid phase synthesis of the template strand of **79**. The efficiency of the "arm" synthesis was analyzed by separating the crude deprotected

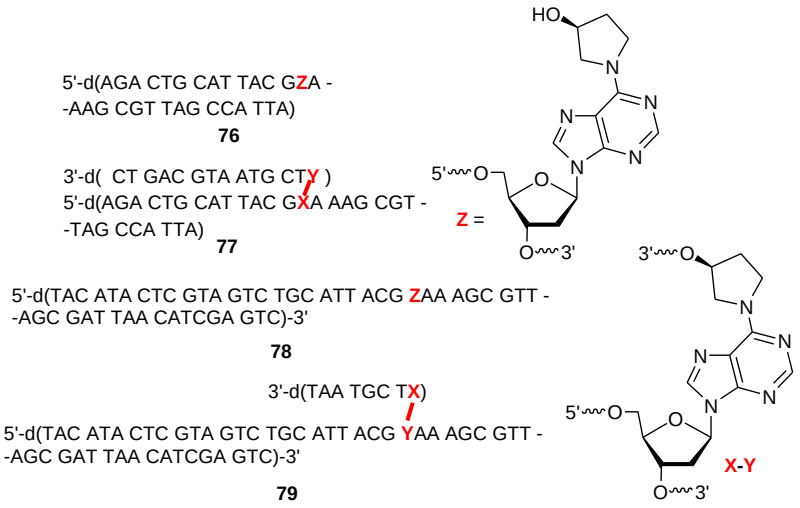


Figure 17. Sequences of template strands **76** and **78** and corresponding DOB ICL mimics **77** and **79**.

oligonucleotide on 20% denaturing PAGE and visualizing the product using a UV lamp and fluorescent TLC plate. In addition to the desired cross-linked oligonucleotide **77**, a significant amount of the corresponding template strand **76** was observed (App. Figure 4). On the other hand, we investigated reverse phase HPLC purification following

synthesis of **79**. Therefore, the DMT protection of the last nucleotide during the synthesis of the opposing strand or "arm" of **79** was left intact. The chromatogram of the synthesis of **79** contained several peaks, all of which were collected (App. Figure 5A). Denaturing PAGE (15%) analysis of the radiolabeled fractions indicated the material in fraction 3 as

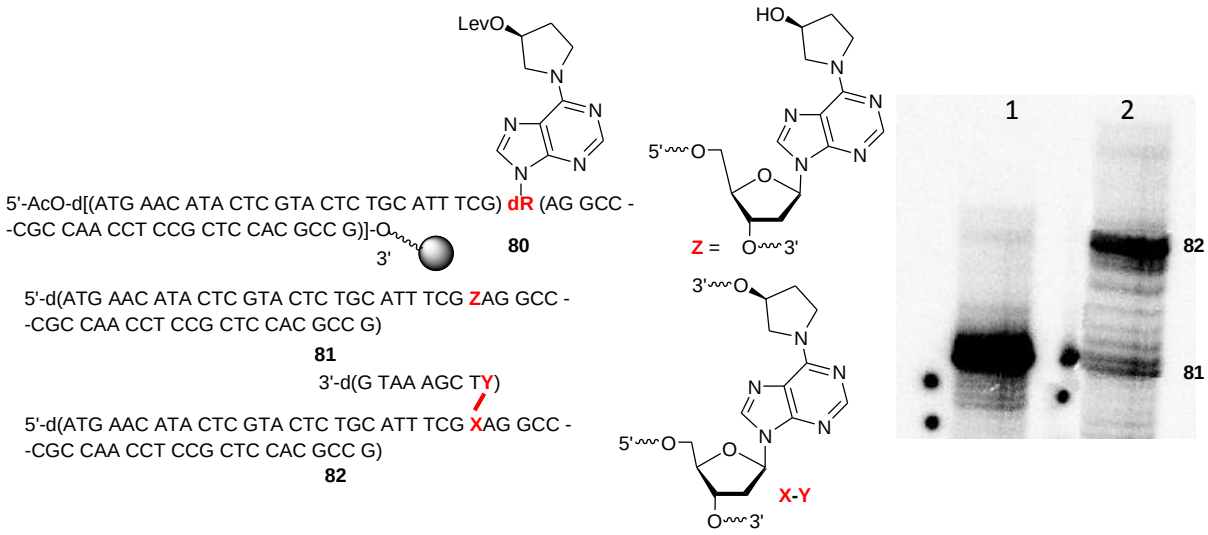


Figure 18. Assessment of delevulinylation efficiency and subsequent extension capability for resin bound and 5'-O-acetyl capped oligonucleotide **80**. Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled, and analyzed on 20% denaturing PAGE. Positions of template strand **81** and corresponding cross-link product **82** on the gel are annotated. Lane 1, **80** deprotected and ^{32}P-labeled. Lane 2, **80** delevulinylated (0.25 M hydrazine in 3:2 pyridine/acetic acid for 25 min, room temperature), extended using 5'-CE "reverse" phosphoramidites, deprotected, and ^{32}P-labeled.

the possible product (**79**), which was later confirmed by MALDI-TOF MS (App. Figure 104). Migration of fraction 2 near the 50 bp DNA from the ladder suggested that it was the template strand (**78**). The large peak area of fraction 2 suggested the presence of a significant amount of the unmodified template strand. In principle, the presence of a unmodified template strand could occur from incomplete delevulinylation or inefficient extension. Both could occur as a consequence of greater load on the solid support due to the longer length of **77** and **79**. In addition, slower moving and (possibly) higher molecular weight products were detected in both cases (uncharacterized top band in App.

Figure 4 and fraction 4 in App. Figure 5A-B). We speculated that the slower moving bands were branched products originating from inefficient capping due to greater load on the resin. Oligonucleotides **76** and **77** were characterized by MALDI TOF MS and ESI-MS (App. Figure 102 and 103).

After speculating about the reason behind incomplete delevulinylation, extension and the origin of branching, we attempted to improve the syntheses by increasing the capping and delevulinylation times. To ensure a more efficient capping, the resin was incubated 25 times (as opposed to 10 times for oligonucleotides ≤ 25 nucleotides) with the capping reagents for 4 min each cycle. In addition, for longer oligonucleotides (≥ 25

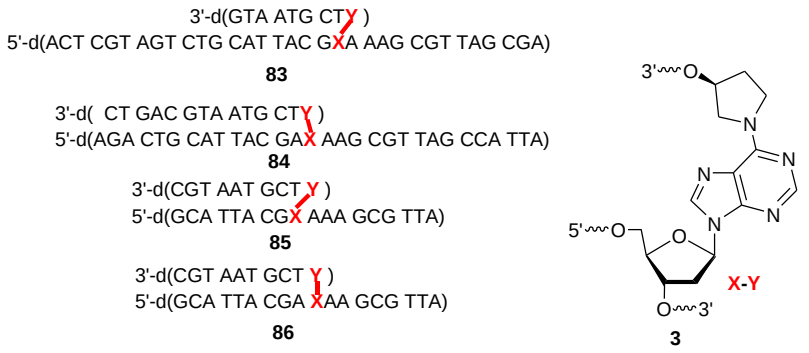


Figure 19. Chemically synthesized cross-linked oligonucleotides containing DOB ICL mimic (**3**).

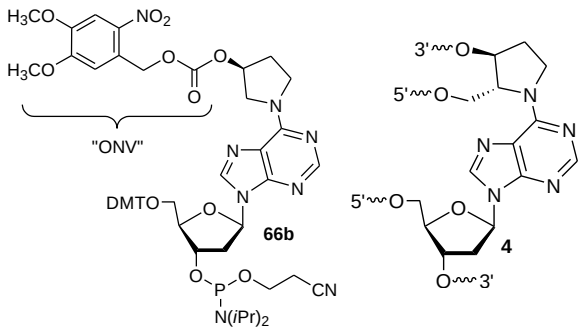
nucleotides), the delevulinylation time was extended to 25 min. To ensure complete reaction of delevulinylated support bound oligonucleotide (**68**), the first phosphoramidite was double coupled (15 min coupling time) while the rest were single coupled for the same duration. The modified procedures were applied for the the synthesis of **82** from resin bound **80**, where the oligonucleotide was subsequently deprotected and cleaved from the support by ammonolysis. Denaturing PAGE (15%) analysis of the crude material following 5'-^{32}P-labeling indicated efficient conversion (87%) of the template strand **81** into **82** without formation of any branched product (Figure 18). Cross-link **82**

was characterized using ESI-MS (App. Figure 105). Using the same procedures, other cross-linked oligonucleotides (**83** – **86**) containing **3** were synthesized (Figure 19), purified by denaturing PAGE, and characterized by either MALDI-TOF MS or ESI-MS (App. Figure 106 – 109).

3.1.1.3. Synthesis of DOB ICL from NVOC phosphoramidite

3.1.1.3.1. Incorporation of photolabile phosphoramidite 66b and photodecomposition of *N4*-benzoyl-dC

After establishing a strategy for DOB cross-link mimic (**3**) synthesis using the levulinyl group, we looked into the employability of photolabile **66b** in ICL synthesis



(Scheme 28). We envisioned that development of *o*-nitroveratryloxycarbonyl (ONV) chemistry along with levulinyl could potentially be helpful in the synthesis of **4**, which requires two orthogonal protections of the 5'- and 3'- hydroxyl group on the cross-linked pyrrolidinyl moiety. Toluene as the photolysis solvent and 2 h irradiation time were adopted based on earlier reports.[130,131] The photolysis tube containing the resin in toluene was equipped with a pipe cleaner as a stirrer to prevent crushing the resin. The synthesis of cross-link **89** was therefore attempted by irradiating (Rayonet photoreactor (model RPR-100) fitted with 16 RPR-3500 lamps having an output maximum (intensity 7 – 9

mW/cm^2) at 350 nm) 5'-capped support bound **87** (Figure 20) for 2 h, followed by extension of the opposing strand ("arm") using regular T phosphoramidites, deprotection

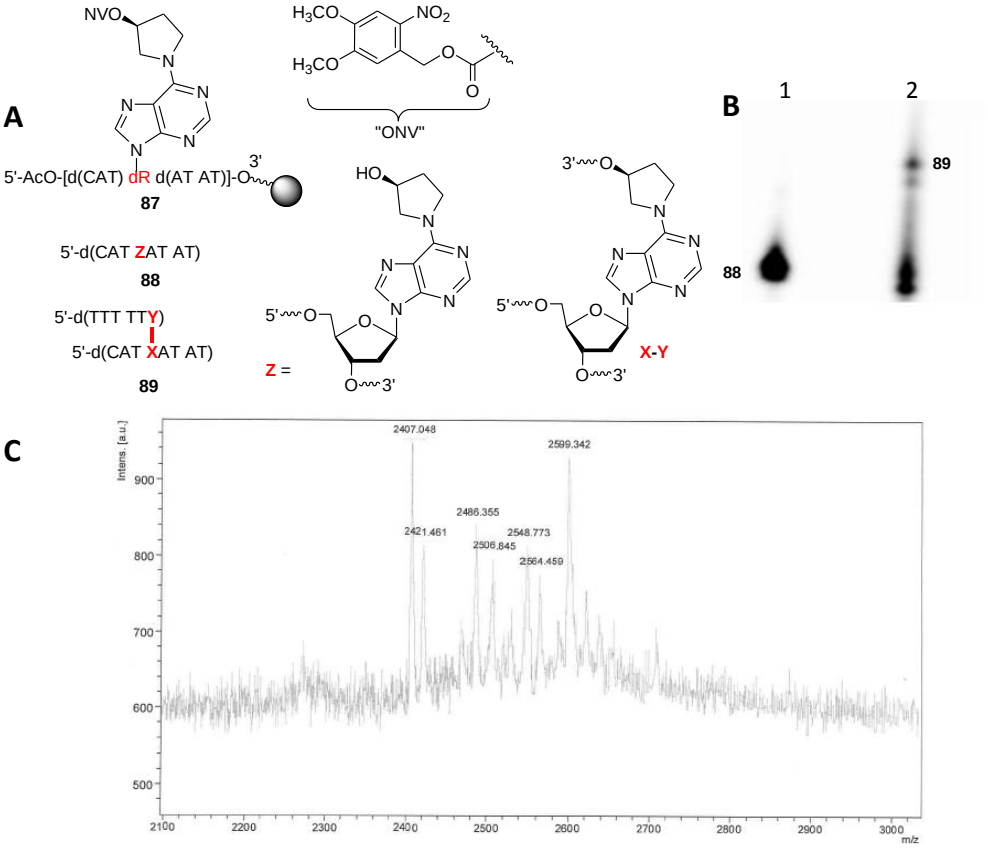


Figure 20. Assessment of photolysis and subsequent extension capability for resin bound and 5'-O-acetyl capped oligonucleotide **87**. (A) Resin bound 5'-O-acetyl capped oligonucleotide **87**, template strand **88** and corresponding DOB ICL mimic **89**. (B) Photolysis and extension of **87** Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Positions of template strand **88** (from unmodified **87**) and corresponding cross-link product **89** on the gel are annotated. Lane 1, **87** deprotected, and 5'-^{32}P-labeled. Lane 2, **87** photolyzed in toluene for 2 h, extended using regular T phosphoramidites, deprotected, and 5'-^{32}P-labeled. (C) MALDI-TOF-MS of recovered photodecomposed "template strand" from lane 2 of panel (B). Calc'd mass, 2462.233, observed mass 2486.355 $[M + Na]^+$, 2506.845 $[M + 2Na]^+$, 2548.773 $[M + 2Na + K]^+$, 2564.459 $[M + 2K]^+$, 2599.342 $[M + 3K]^+$.

and cleavage from the resin by ammonolysis. However, analysis of the 5'-^{32}P-labeled crude material on 20% denaturing PAGE and MALDI-TOF MS indicated a modest yield (18%) of **89** (App. Figure 110) along with decomposition of template strand **88** (Figure 20). The cause of the decomposition was traced to the previously reported susceptibility of N4-benzoyl dC protection to photodecomposition under UV irradiation.[131,132]

3.1.1.3.2. Replacement of *N6*-benzoyl-A and *N4*-benzoyl-C by isobutyryl-C and isobutyryl A for template strand synthesis

To assess the vulnerability of *N*-benzoyl protecting groups to UV radiation, we measured the absorbances of *N*6-benzoyl-A and *N*4-benzoyl-C phosphoramidites (54 μM) between 350 and 400 nm and compared them with that of **63b** (Figure 21 and App. Figure 6). The spectral distribution graph of RPR-3500 lamp used in Rayonet photreactor (http://www.rayonet.org/graphscharts.htm) suggests the wavelengths of the light emitted

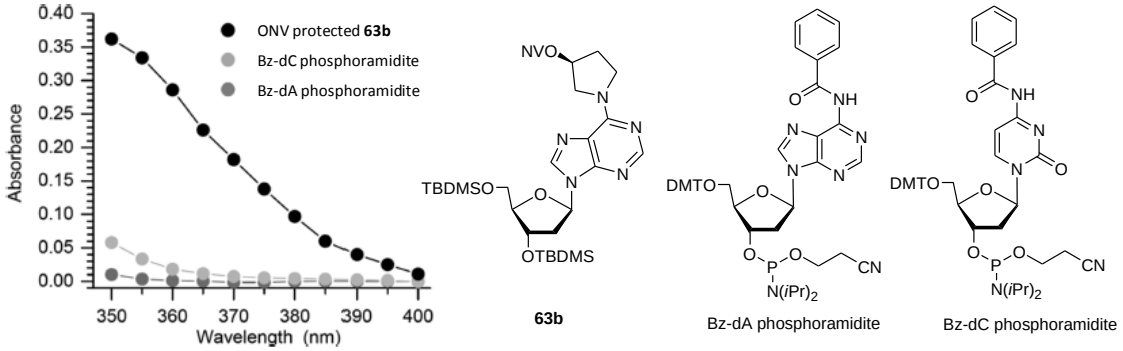


Figure 21. Absorbance of nitroveratryloxycarbonyl protected **63b**, and benzoyl protected A and C phosphoramidites between 350 and 400 nm. Concentrations of all solutions were 54 μM.

from the lamp spans over the range 300 – 400 nm, with the maximum emission occurring at 350 nm. While the λ_{max} of **63b** was at 350 nm (Figure 21), the same for *N4*-benzoyl-C and *N*6-benzoyl-A phosphoramidites were at 315 and 330 nm, making them susceptible to photodamage (App. Figure 6). Instead of changing the light source from 350 nm, we explored the possibility of using alternative protecting groups that do not absorb light at 350 nm. Attractive possibilities were commercially available *N6*-isobutyryl-A and *N4*-isobutyryl-C phosphoramidites. As we intended to use both ONV and levulinyl protecting groups during the C4-AP ICL analogue (**4**) synthesis, isobutyryl (iBu) had to be compatible with delevulinylation too. This was addressed by making oligonucleotides containing a single iBu-A (**90**) or iBu-C (**93**), exposing them to delevulinylation

treatment, and subjecting them to solid phase synthesis (Figure 22). After ammonolysis of the synthesized oligonucleotides, an aliquot from the crude was 5'-^{32}P-labeled, and analyzed on 20% denaturing PAGE gel (App. Figure 7A and B). If the iBu protections withstand delevulinylation, only the template strands (**91** and **94**) will be visible on the

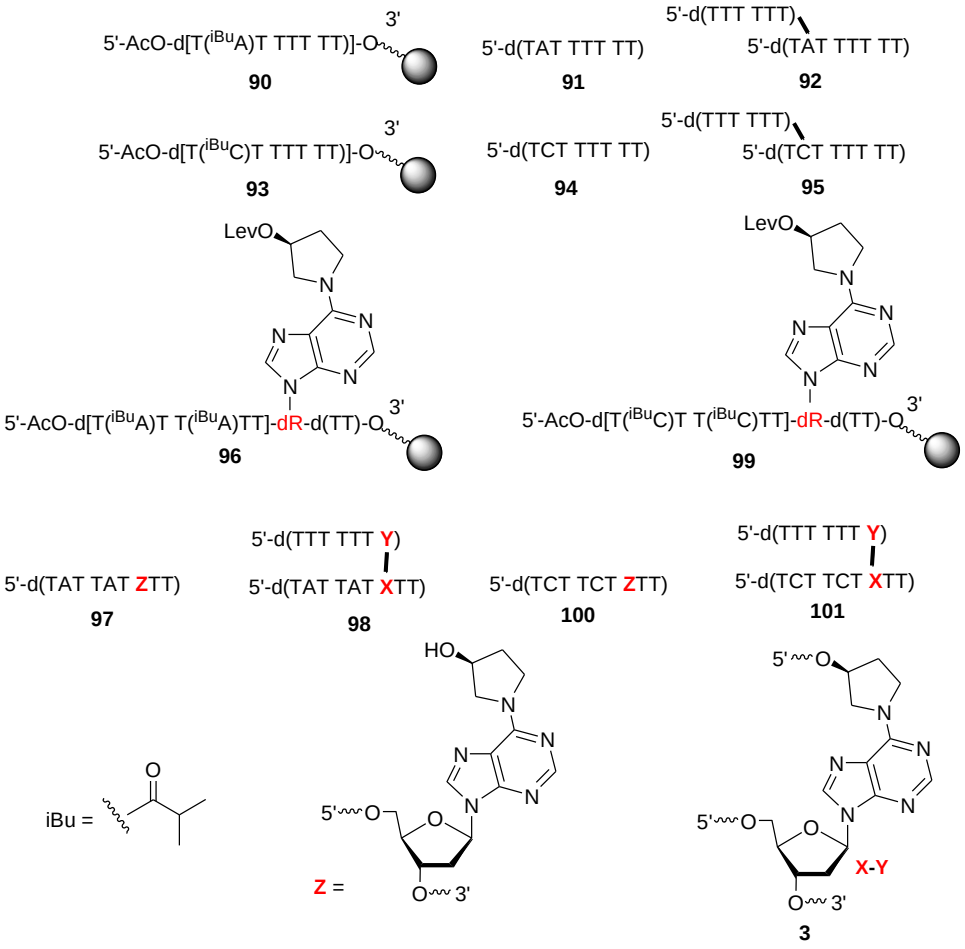


Figure 22. Sequences of support bound 5'-capped template strands **90**, **93**, **96**, and **99**, deprotected template strands **91**, **94**, **97** and **100**, branching products **92** and **95**, and DOB ICL mimics **98** and **101**.

gel. However, if the groups do not survive, branched oligonucleotides (**92** and **95**) will be formed. To distinguish between the extension products resulting from inefficient capping and those from branching (**92** and **95**), we subjected the capped **90** and **93** to the same solid phase extension without delevulinylation (Lane 2, App. Figure 7A and B, Table 2). Although **90**, containing a single iBu-A survived the delevulinylation (extension yield

2.84 ± 0.08% and 2.76% with and without delevulinylation, respectively, Figure 23A and App. Figure 7A), the presence of a single iBu-dC led to 3.32 ± 1.01% extension for delevulinylated **93** (compared to 0.1% extension without delevulinylation, Figure 23B and App. Figure 7B).

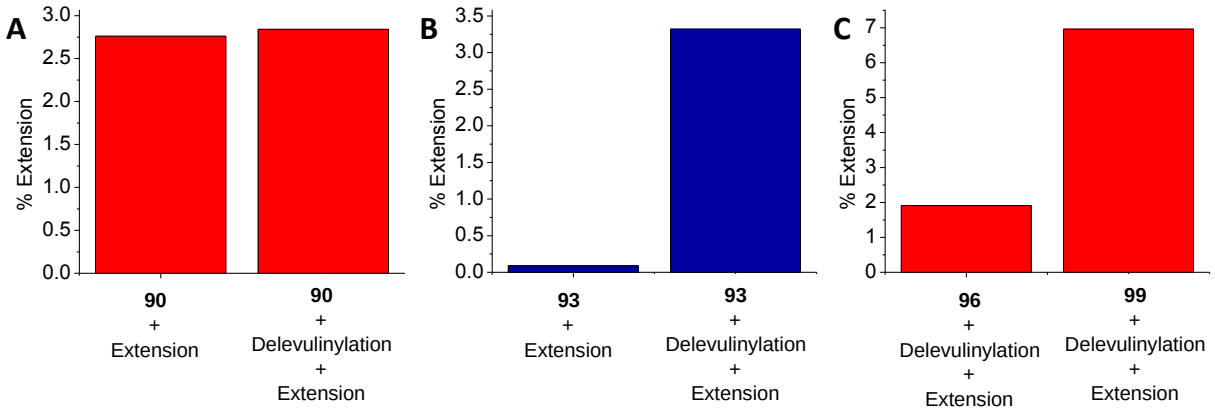


Figure 23. Extension of oligonucleotides **90**, **93**, **96** and **99** after delevulinylation. (A) Comparison of extension of **90** with or without delevulinylation. (B) Comparison of extension of **93** with or without delevulinylation. (C) Comparison of extension of **96** and **99** after delevulinylation. Extension yields were calculated based on the band intensity of template strand and final product.

This was further validated during the synthesis of the DOB ICLs from resin bound oligonucleotides **96** and **99** (Figure 22) with two iBu-dAs and iBu-Cs, respectively, and a levulinyl protected DOB ICL mimic. We probed their delevulinylation and subsequent solid phase extension to produce **98** and **101** by deprotection, ^{32}P-labeling and gel analysis as described before (App. Figure 7). Reaffirming our effective capping and delevulinylation strategies, no unwanted extension without delevulinylation (branching products) or leftover template strand (**97** and **100**, from unmodified **96** and **99**) was detected (Lane 2, 3, 4 and 5, App Figure 7C). In accordance with the previous finding on iBu-A and iBu-C, the latter caused greater yield of unwanted extension product (1.91 and 6.16% for iBu-A and iBu-C, respectively,

Figure 23C, App. Figure 7C). Therefore, the strategy to replace the photolabile *N6*-benzoyl-A and *N4*-benzoyl-C with their isobutyryl protected analogues was abandoned.

3.1.1.3.3. Synthesis of DOB ICL from nitroveratryloxycarbonyl phosphoramidite using a long-pass optical filter

A reexamination of absorbance of **66b** and benzoyl protected A and C phosphoramidites (Figure 21) revealed that the absorption of the benzoyl group is negligible above 375 nm. In contrast, the *o*-nitroveratryl group still has significant absorption at this wavelength. Consequently, we used a 375 nm long pass filter (Newport no. CGA-375, App. Figure 8) that had zero transmittance at 370 nm and 50% at 375 nm (App. Figure 8). The filter was fitted into a Schoeffel 1000 W short-arc lamp (model LH151N/2, intensity 9 mW/cm^2 at 350 nm). When the ONV protected support bound oligonucleotide **87** was deprotected via the filter-aided-photolysis (2 h in toluene, room temperature) and subjected to solid phase extension, cross-link **89** was formed in a modest 30% yield (App. Figure 9). In contrast to the previous photolysis without filter, template strand **88** (produced from unmodified **87**) did not divulge any sign of decomposition of the benzoyl groups. Having overcome the vulnerability of *N*4-benzoyl-C and *N*6-benzoyl-A from photolysis, we focused on improving the extension yield.

When support bound *o*-nitroveratryloxycarbonyl (NV) protected **87** was converted to **89** by photolysis in toluene for 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, and 4.0 h (room temperature), followed by extension, the highest yields were obtained at 1 h (35.8%) and 2.5 h (32.5%) (App. Figure 10). The yield dropped after 1 h, surged at 2.5 h, and then continued to drop, reaching 16% at 4 h. While the gradual drop after 2.5 or 3 h could be attributable to photodecomposition of **87** due to prolonged photo-exposure, we could not readily explain the low yields at 1.5 or 2.0 h. When the same support bound 5'-capped

oligonucleotide (**87**) was photolyzed in 9:1 acetonitrile/water (2 h, room temperature) prior to solid phase synthesis, it produced 50% extension (App. Figure 11). We expected that the hydrolysis of the photoreduced NV group will be easier in an acidic medium. In accordance with our hypothesis, 90% extension yield was obtained when the solid support was washed with 4:1 pyridine/acetic acid (pH 5.5) for 10 min at room temperature following photolysis in 9:1 acetonitrile/water (App. Figure 11). The absence

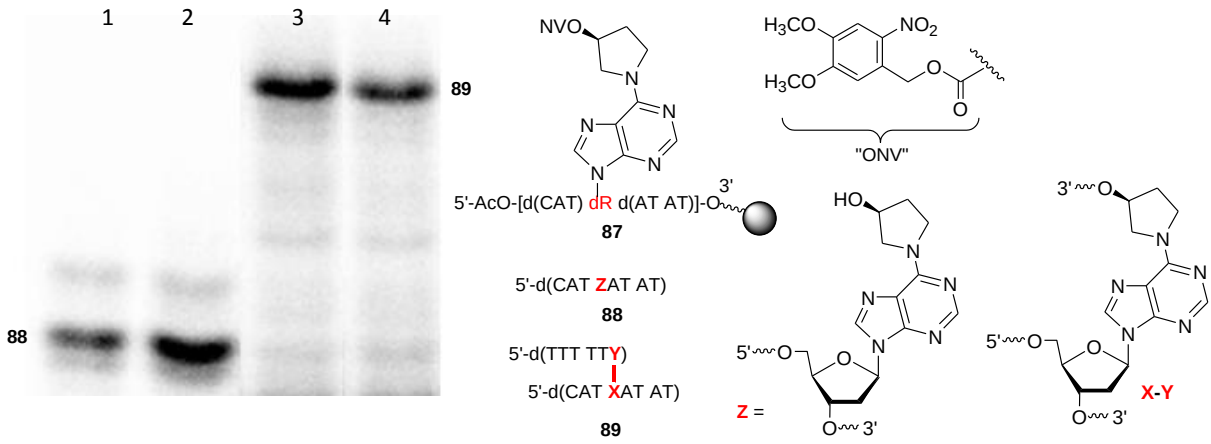


Figure 24. Assessment of filter-aided-photolysis (375 nm long pass) and subsequent extension capability for resin bound and 5'-O-acetyl capped oligonucleotide **87**. Oligonucleotide was subsequently deprotected and cleaved from the support by ammonolysis. An aliquot from the crude was 5'-^{32}P-radiolabeled, and loaded on 20% denaturing PAGE gel. Positions of the template strand **88** (from unmodified **87**) and corresponding cross-link product **89** on the gel are annotated. Lane 1, **87** deprotected without any modification. Lane 2, **87** subjected to extension using regular T phosphoramidite. Lane 3 and 4, **87** subjected to 2.0 and 2.5 h of photolysis followed by extension using regular T phosphoramidites, respectively.

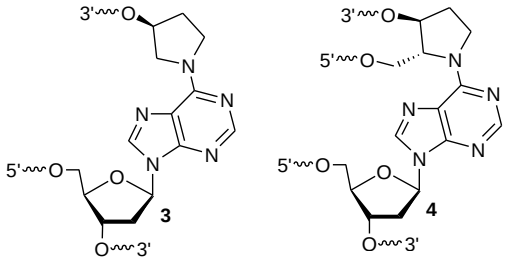
of any extension for support bound **87** after washing it under the same conditions confirmed that the removal of NV group was not due to the treatment with the 4:1 pyridine/acetic acid mixture (App. Figure 11). However, attempts to replicate this astonishing extension yield by washing with 4:1 pyridine/acetic acid was unsuccessful irrespective of the solvent used in photolysis (toluene or 9:1 acetonitrile/water, App. Figure 12). Instead, we noticed an unexplained decomposition of **88** when photolysis was followed by 4:1 pyridine/acetic acid wash (10 min, room temperature). Despite the

setback in the pyridine/acetic washing procedure, we were finally able to efficiently convert **87** into **89** by a simple mechanical readjustment of the light source that enabled the beam to cover the whole solution for more efficient irradiation (Figure 24). Solid phase extension following photolysis for 2.0 and 2.5 h provided nearly quantitative conversion of **87** to **89**. We adopted these photolysis durations (2.0 and 2.5 h for template strand length shorter or longer than 25 nucleotides, respectively) in toluene for use in the C4-AP cross-link mimic (**4**) synthesis.

3.1.2. Synthesis of C4-AP ICL mimic

3.1.2.1. Synthesis of C4-AP ICL mimic from 5'-ONV-3'-Lev-C4-AP phosphoramidite

3.1.2.1.1. Removal of *o*-nitroveratryloxycarbonyl and levulinyl protecting groups from 5'- and 3'- hydroxyls of pyrrolidine



After establishing the compatibility of the levulinyl (Lev) and *o*-nitroveratryloxycarbonyl (NV) groups with solid phase synthesis conditions, we sought to synthesize C4-AP ICL mimic **4** using them. We decided to implement the solid phase synthesis and capping conditions developed for the DOB analogue for this purpose. In addition, we explored combining allyloxycarbonyl (ALLOC) and levulinyl for synthesizing **4**. We envisioned that the two hydroxyls in the pyrrolidinol ring in **4** would be protected with NV (or ALLOC) and levulinyl. We attempted to determine which of

them would protect the 5'- and 3'- hydroxyls of pyrrolidinol and more importantly, which would be removed first during the course of the solid phase synthesis. To determine whether NV or levulinyl could be removed first, we subjected support bound 5'-capped levulinyl protected oligonucleotide **102** to photolysis in toluene (2 h, room temperature), followed by delevulinylation (0.25 M hydrazine, 3:2 pyridine/acetic acid, 10 min, room temperature) and solid phase extension (Figure 25 and App. Figure 13). Based on the strategy established for DOB ICL analogue **3**, all photoirradiation experiments were carried out using a Schoeffel 1000 W short-arc lamp (model LH151N/2, intensity 9 mW/cm^2 at 350 nm) fitted with a 375 nm long pass filter. The product or extension efficiency did not differ from that done without photolysis (Lane 3 and 4, App. Figure 13). Having established that photolysis followed by hydrazine treatment did not affect extension, we sought to explore the consequences when the order of steps were reversed. The support bound NV protected 5'-capped-**87** was subjected to delevulinylation (0.25 M hydrazine in 3:2 pyridine/acetic acid, 10 min, room

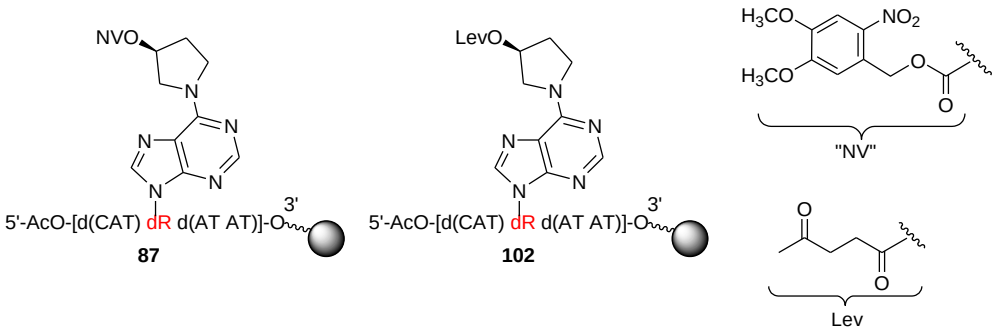
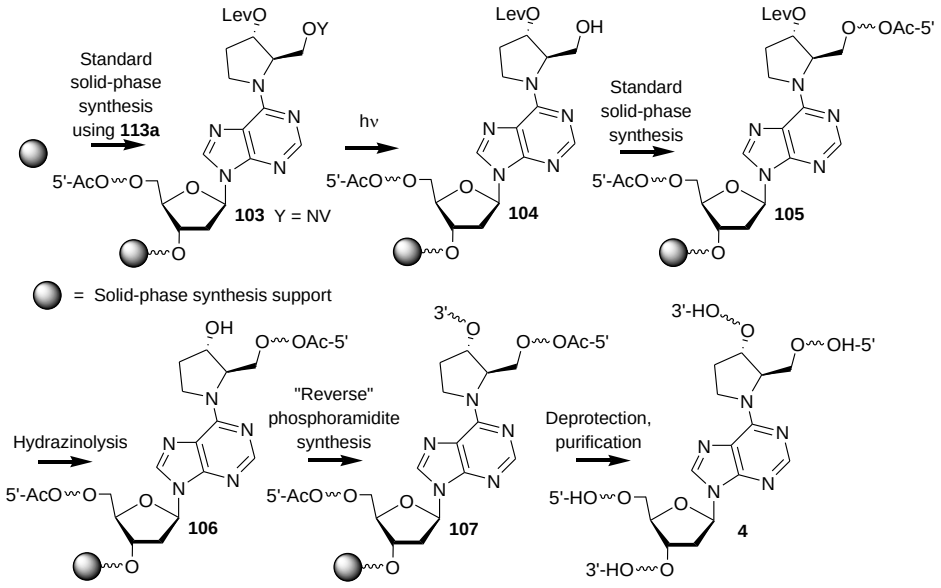


Figure 25. Support bound oligonucleotides **87** and **102**.

temperature) prior to subsequent photolysis (toluene, 2 h, room temperature) and solid phase extension. While only photolysis followed by extension successfully yielded cross-link **89** (Lane 5, App. Figure 14), delevulinylation preceding the photoirradiation led to

decomposition (Lane 4, App. Figure 14). This was consistent with the previously observed decomposition of oligonucleotide for photolysis followed by pyridine/acetic acid washing. Although unexplainable, the observation led us to abandon the second route, delevulinylation and 3'-arm synthesis followed by photolysis and 5'-arm synthesis. Next, we sought to resolve the preferential location of NV and levulinyl group on 5' and 3' hydroxyls of the pyrrolidine ring of C4-AP ICL analogue. Similar to DOB ICL mimic analogue **3**, the solid phase synthesis of the 3'-arm involves less efficient 3' to 5' coupling using 5'-cyanoethyl (CE) reverse phosphoramidites. Therefore, we decided to employ the levulinyl group to protect the 3'-hydroxyl of the pyrrolidinol, anticipating its efficient removal compared to the NV (as seen during synthesis of DOB ICL analogue **3**) would compensate for the relatively inefficient 3' to 5' extension. Hence,

Scheme 29. Synthesis of C4-AP ICL mimic 4 by orthogonal deprotection strategy



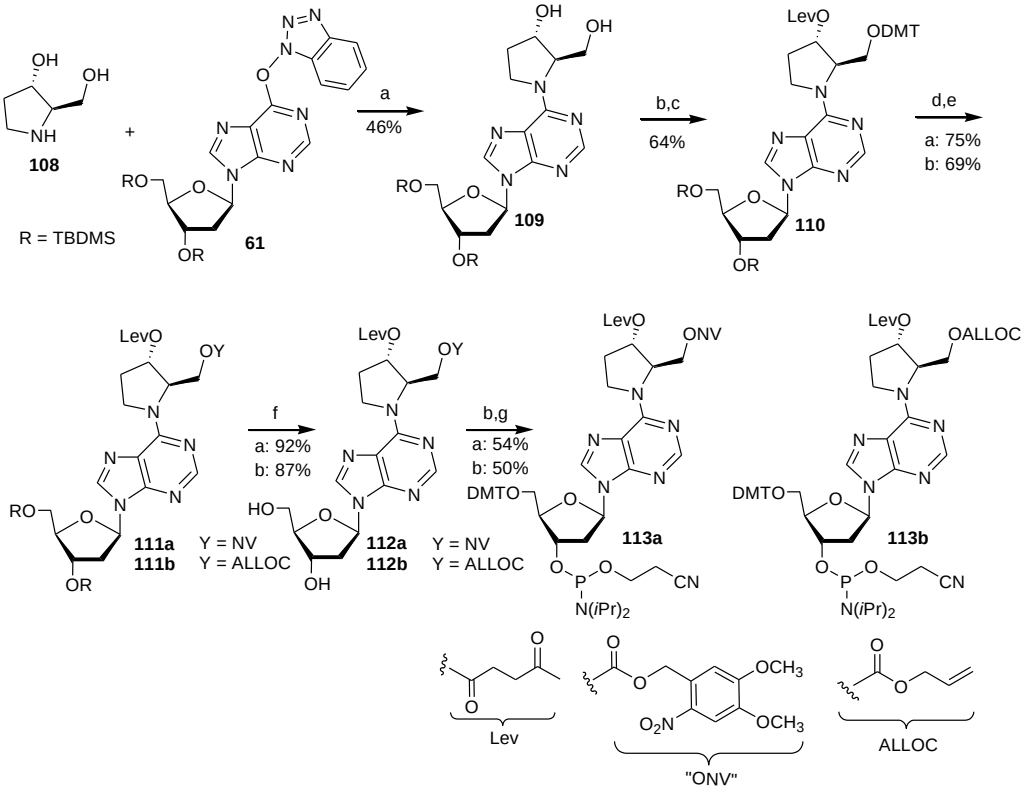
the 5'-hydroxyl was protected with the NV group, so that its inefficient removal could be balanced by the more efficient 5' to 3' synthesis using 5'-dimethoxytrityl (DMT) 3'-CE "regular" phosphoramidites (Scheme 29). We used the same argument for making a C4-

AP ICL analogue using ALLOC and levulinyl groups, putting the ALLOC group on the 5'-OH of the pyrrolidinol moiety of **4**.

According to the strategy of synthesizing **4** (Scheme 29), requisite phosphoramidite (**113a**) would be employed to make the template strand (**103**). Following acetyl capping of the template strand, the NV group would be removed from **103** (Scheme 29). After extending the primary alcohol of **104** using "regular" phosphoramidites, the nascent 5'-arm would be acetyl capped (**105**). Following hydrazinolysis (0.25 M hydrazine in 3:2 pyridine/acetic acid, 10 or 25 min, room temperature) of **105**, the free 3'-hydroxyl of **106** would be extended using 5'-CE "reverse" phosphoramidites (**107**). Finally, ammonolysis and cleavage of **107** containing both 5' and 3' arms from the solid phase support, purification by denaturing PAGE would yield **4**. We anticipated that deprotection of levulinyl and NV groups and subsequent extensions would follow methodologies developed earlier during the synthesis of **3**.

The synthetic strategy (Scheme 30) for making the requisite C4-AP ICL mimic phosphoramidite (**113a**) was derived from methodology developed for the synthesis of DOB ICL mimic phosphoramidites. The same strategy was implemented to make the C4-AP ICL mimic phosphoramidite containing the ALLOC protecting group (**113b**). Similar S_NAr chemistry between **108** and **61** led to production of cross-link analogue **109**. We could not selectively protect the 5'-pyrrolidinyl hydroxyl with a nitroveratryloxycarbonyl (NV) or allyloxycarbonyl (ALLOC) moiety by treating **109** with the corresponding chloroformates. Therefore, we protected it with DMT followed by functionalization of

Scheme 30. Synthesis of C4-AP ICL mimic phosphoramidites 113a,b

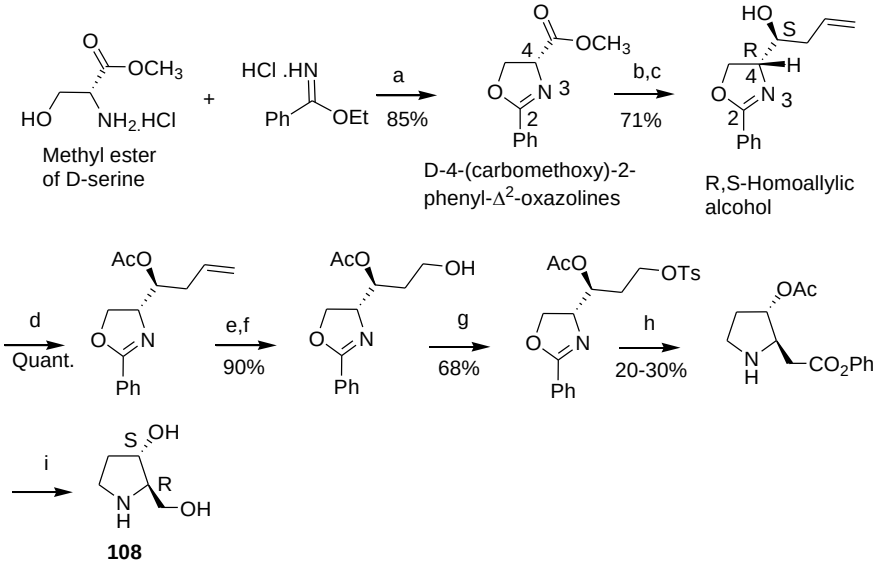


Key: a) Cs_2CO_3, THF b) DMTCl, pyridine c) Levulinic acid, EDCI, DMAP, DMF
d) AcOH, H_2O, THF (3:1:1) e) *o*-nitroveratryl or allyloxycarbonyl chloroformate, DMAP, CH_2Cl_2
c) Et_3N•3HF, THF d) DMTCl, pyridine e) 2-Cyanoethyl chlorophosphoramidite, diisopropylethylamine, CH_2Cl_2

the 3'-pyrrolidinyl hydroxyl with the levulinyl group (**110**). Afterwards, DMT protecting group on 5'-pyrrolidinyl OH was removed to have it protected with NV (**111a**) or ALLOC (**111b**). Following desilylation, **112a** and **112b** were converted to the corresponding phosphoramidites (**113a**,**b**) using standard procedures.

The first attempt at synthesizing of (*R*,*S*)-3-hydroxy-2-hydroxymethylpyrrolidine (**108**) was done using the strategy reported in Scheme 31.[133] Starting from methyl ester of D-serine, this method involves synthesis of an oxazoline ring which later rearranges to the desired pyrrolidine framework. The stereochemistry at C4-position of the final pyrrolidine ring was derived from the chirality of serine itself. The stereochemistry of the

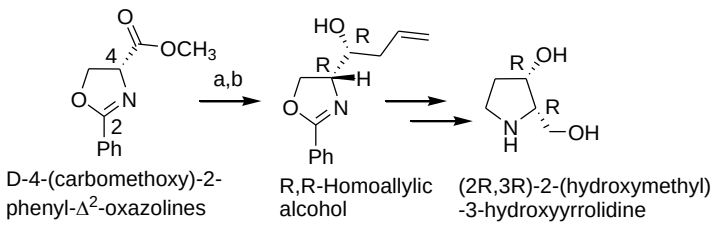
Scheme 31. Synthesis of 108



Key: a) DIPEA, DCM, reflux; b) DIBAL-H, MeOH quench; c) (+)-$IPc_2B(CH_2CH=CH_2)$, -78 ^oC d) Ac_2O, triethylamine; e)OsO_4, THF, $NaIO_4$; f) $NaBH_4$, MeOH; g) NaH, p-toluenesulfonyl chloride; h) 1 M HCl in MeOH; i) 5 M NaOH in MeOH;

hydroxyl substitution of the R,S-homoallylic alcohol was the result of a diastereoselective nucleophilic attack by (+)-$IPc_2B(CH_2CH=CH_2)$ on the carbonyl (Scheme 31). Scheme 32, on the other hand is an almost identical synthetic sequence described in the same report[133] to make (*R*,*R*)-3-hydroxy-2-hydroxymethylpyrrolidine. It differs from Scheme 31 only in the direction of diastereoselective nucleophilic attack at the second step using (–)-$IPc_2B(CH_2CH=CH_2)$ instead of (+)-$IPc_2B(CH_2CH=CH_2)$. The changed directionality of the nucleophilic addition thus changes the stereochemistry (S to R) of the hydroxyl

Scheme 32. Synthesis of (2R,3R)-2-(hydroxymethyl)-3-hydroxyyrrolidine



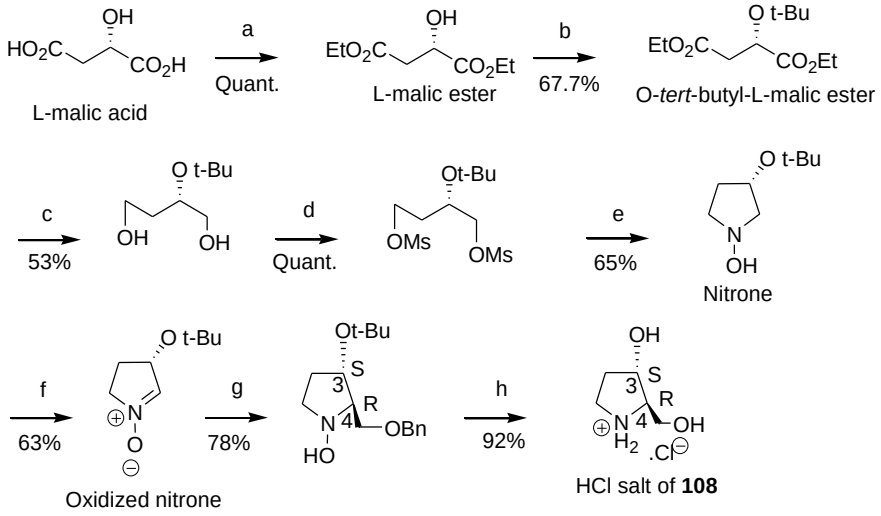
Key: a) DIBAL-H, MeOH quench; b) (-)-$IPc_2B(CH_2CH=CH_2)$, -78 ^oC.

substitution of the resultant homoallylic alcohol (Scheme 32). During our attempts to synthesize **108** (Scheme 31), the NMR and specific rotation of the prepared R,S-homoallylic alcohol matched with that of the R,R-homoallylic alcohol reported in the article.[133] NOE studies to investigate the stereochemical relationship between C4-H of the homoallylic alcohol and the hydrogen α to OH were unsuccessful due to rotation of the connecting C-C bond. Assuming the NMR mix-up as an inadvertent mistake of the authors, we continued the synthesis, anticipating to resolve the ambiguity at the final step (Scheme 31). The rearrangement of oxazoline to pyrrolidine at the penultimate step was successful, albeit in low yield. We successfully accomplished the final deprotection step using conditions as described (Scheme 31).[133] Upon completion, we discovered that the NMR (of the final molecule made by us and that reported in the article[133]) did not match with (*R*,*S*)-3-hydroxy-2-hydroxymethylpyrrolidine **(108**) synthesized by other independent strategies.[134,135] The article from which Scheme 31 was adopted did not attempt to verify the final molecule with a crystal structure.[133] Since all other reactions in Scheme 31 were well-established, we suspected that the key rearrangement at the penultimate step did not yield the pyrrolidine ring as described. If our suspicion was correct, the final molecule in Scheme 31 may not be **108**. We did not investigate this approach any further, and used a different strategy to make **108**.

Our second attempt to make the hydrochloride salt of **108** exploited the chirality of the α-hydroxyl group in L-malic acid (Scheme 33).[136,137] The advantage of this strategy lies in the fact that a number of steps can be carried out without chromatographic purification, which makes the approach considerably more efficient than the previous

one. In addition, it helped us when scaling up the synthesis. In this scheme, the α-hydroxyl group of L-malic ester was efficiently protected with a *tert*-butyl group to

Scheme 33. Synthesis of hydrochloride salt of 108



Key: a) Ethanol, conc H_2SO_4, reflux; b) isobutene, H_3PO_4, P_2O_5, $BF_3.Et_2O$, THF; c) $LiAlH_4$, reflux; d) MsCl, Et_3N, DCM; e) $NH_2OH.HCl$, Et_3N, reflux; f) HgO, DCM; g) Bu_3SnCH_2OBn, nBuLi, THF; h) $Pd(OH)_2$-C, H_2, HCl in MeOH.

generate an O-*tert*-butyl protected ester.[136] The stereochemistry of the protected alcohol was retained in the C3 hydroxyl of **108**. Later in this strategy, the oxidized nitrone was subjected to nucleophilic attack by Bu_3SnCH_2OBn. The trans stereoselectivity (trans/cis > 95:5) of the nucleophilic attack with respect to the O-*tert*-butyl group generated the desired S stereochemistry at C4 position. Finally, removal of benzyl and *tert*-butyl protecting groups (by acid treatment and hydrogenolysis, respectively) produced the hydrochloride salt of **108** (Scheme 33).[137]

3.1.2.1.2. Synthesis of cross-linked oligonucleotides 117 containing C4-AP ICL analogue 4

The synthesis of a short (8 bp, **117**, Figure 26) oligonucleotide containing C4-AP ICL analogue **4** was attempted as a proof of concept. The phosphoramidite **113a** was

incorporated into the template strand by manual coupling (15 min) with 80% yield, while the average stepwise yield was 90.3%. During the synthesis of the template strand, commercial 5'-DMT 3'-CE "regular" phosphoramidites (benzoyl-A, isobutyryl-G, and benzoyl C) were employed. Following 5'-acetyl capping of the support bound template strand containing functionalized C4-AP analogue (**114**), a portion of the resin (typically

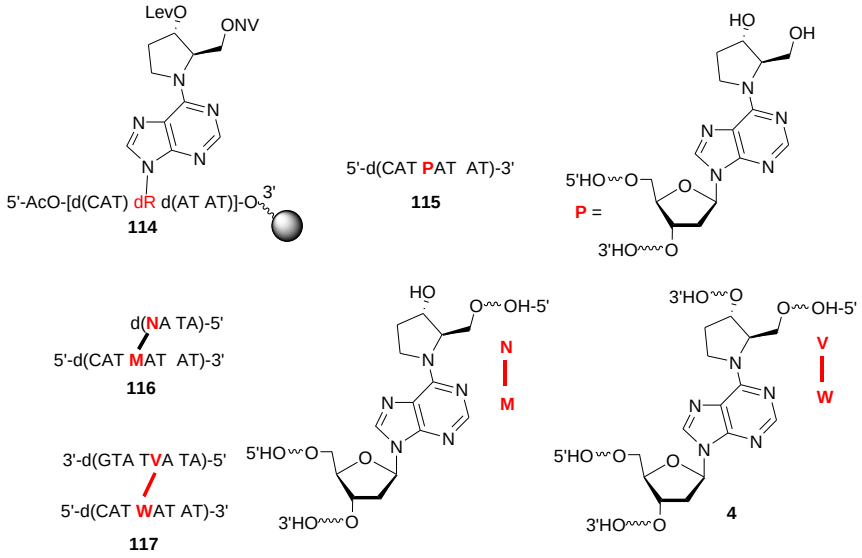


Figure 26. Sequences of resin bound oligonucleotide **114**, template strand **115**, ICL **116** after 5'-extension and **117** after both 5'- and 3'- extension.

1/5th or 1/6th of the total amount of resin present in a 1 μmol scale column) was subjected to photolysis as described before for 1.0, 1.5 and 2.0 h (App. Figure 15). The 5'-arm was synthesized with the first phosphoramidite being manually coupled. The rest of the phosphoramidites (concentration 0.15 M) were incorporated into the arm via automated synthesis cycle (coupling time 15 min) to produce **116**. Following the synthesis of the arm, an aliquot from the crude was ^{32}P-labeled and analyzed on 20% denaturing PAGE. The highest yields were obtained for 1.5 h (78%) and 2.0 h (71.3%) of photolysis (App. Figure 15A and 15B). In addition, template strand **115** remained intact on the gel implying absence of photodecomposition of benzoyl-dC.

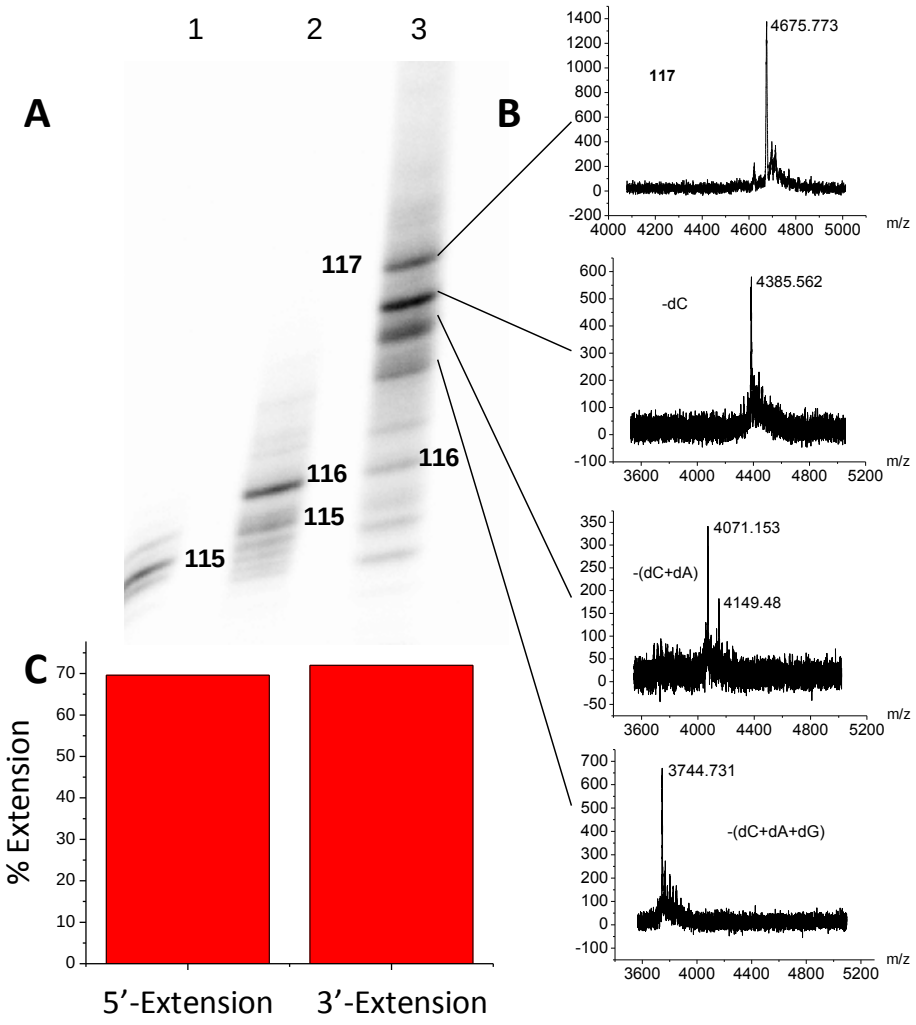


Figure 27. Filter aided photolysis and delevulinylation followed by solid phase extensions of support bound oligonucleotide **114.** Oligonucleotides from each step were subsequently deprotected and cleaved from the support by ammonolysis. An aliquot from the crude was 5'-^{32}P-labeled, and loaded on 20% denaturing PAGE gel. Positions of template strand **115** (from unmodified **114**), ICL **116** and **117** are annotated. (A) Observation of 5'- and 3'-extension on 20% denaturing PAGE gel. Lane 1, template strand **115** (from unmodified **114**). Lane 2, **114** photolyzed for 2 h in toluene and extended using "regular" phosphoramidites. Lane 3, Resin from 5' extension was capped, and subjected to delevulinylation treatment (0.25 M hydrazine in 3:2 pyridine/acetic acid, 10 min, room temperature) and solid phase extension using 5'-CE "reverse" phosphoramidites. (B) MALDI-TOF-MS characterization of **117** and the deletion products. (C) Amount of extension product formed with respect to starting material for 5' and 3'-arm synthesis. All product bands were taken into account and compared with that from starting material during the computation of extension yield.

Next, we attempted the synthesis of full length ICL **117**. After photolysis and synthesis of the 5'- arm, **117** was obtained by delevulinylation and synthesis of the 3'-arm using 5'-cyanoethyl (CE) "reverse" phosphoramidites. The success of the synthesis was

analyzed by ^{32}P-labeling the deprotected crude and analyzed them on 20% denaturing PAGE (Figure 27A). The synthesis yielded 2.7 nmol of ICL **117** from 1/5th of the resin present in a 1 μmol scale column. Although the 5'- and 3'- extensions proceeded with efficient conversion of the starting material into products (70% and 73.5%, respectively, Figure 27C), the gel indicated the presence of a mixture of products (Lane 3, Figure 27A). Upon isolation and characterization using MALDI-TOF MS, the top band on lane 3 was identified as the desired **117** (calc'd mass 4674.083, found 4675.773, Figure 27B). The rest of the bands on the same lane were characterized as deletion products with either missing a C or (C + A) or (C + A + G) (Figure 27B). As the deletion products did not correlate with the sequence of the 3'-arm, we hypothesized their origin to be elsewhere during the ICL synthesis. Since we wanted to synthesize longer C4-AP ICL mimics (template strand length ≥ 25 nucleotides), the presence of the deletion products convinced us to adapt more stringent solid phase synthesis protocols.

3.1.2.1.3. Applicability of chemical phosphorylating agents during synthesis of longer C4-AP ICL mimic (template strand length ≥ 25 nucleotides)

To minimize deletion product generation during the synthesis of ICL **121** (Figure 28), phosphoramidite **113a** was incorporated into the template strand by manual double-coupling (15 min each). Furthermore, 1000 Å resin was employed to withstand the load of longer (≥ 25 nt) length of template strand synthesis. In addition, automated double-coupling of the native phosphoramidites were used during solid phase synthesis of the template strand (**119**). Coupling of phosphoramidite **113a** occurred in a quantitative manner while 98.2% average stepwise yield was obtained. We attempted to phosphorylate the 5'-terminus using solid phase chemical phosphorylation with a

commercially available masked phosphate amidite (**122** or **123**, Figure 29). We expected the presence of a phosphate to aid us during subsequent T4 DNA ligase mediated ligation and selective 5'-^{32}P-labeling to synthesize a full-length cross-link product. Since the synthesis of the 5' arm would be succeeded by delevulinylation (conversion of **120** to

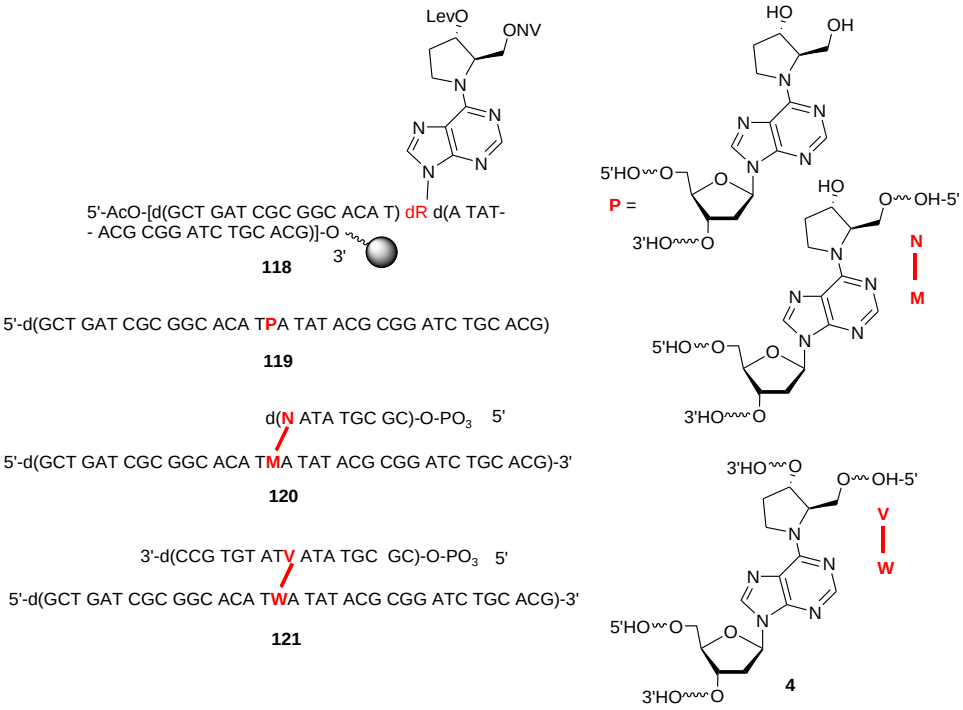


Figure 28. Sequences of support bound oligonucleotide **118**, template strand **119**, ICL **120** after 5'- extension and **121** after both 5'- and 3'- extension.

121, Figure 28), the masking groups on the chemical phosphorylating agents (CONHMe and CO_2Et) needed to survive the hydrazinolysis.

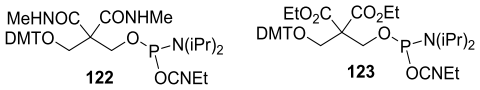
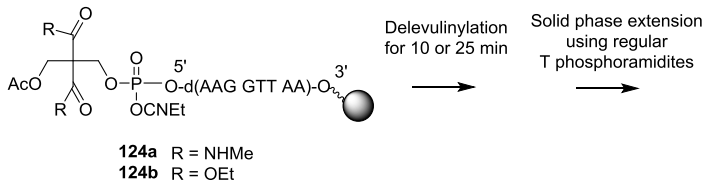


Figure 29. Commercially purchased chemical phosphorylating reagents.

To assess the compatibility of **122** and **123** with delevulinylation, we synthesized oligonucleotides containing the masked phosphates (**124a**,**b**, Scheme 34). After 5'-acyl protection, the oligonucleotides were subjected to delevulinylation (0.25 M hydrazine in

3:2 pyridine/acetic acid, room temperature) for 10 or 25 min, followed by solid phase extension using regular T phosphoramidites. The results were analyzed by deprotecting

Scheme 34: Determination of compatibility of phosphorylating reagents with delevulinylation



the oligonucleotides from support, ^{32}P-labeling the crude and analyzing them on 20% denaturing PAGE. If the protecting groups on the phosphorylating agents are stable to hydrazinolysis, no extension would be detected on the gel. Both groups were somewhat

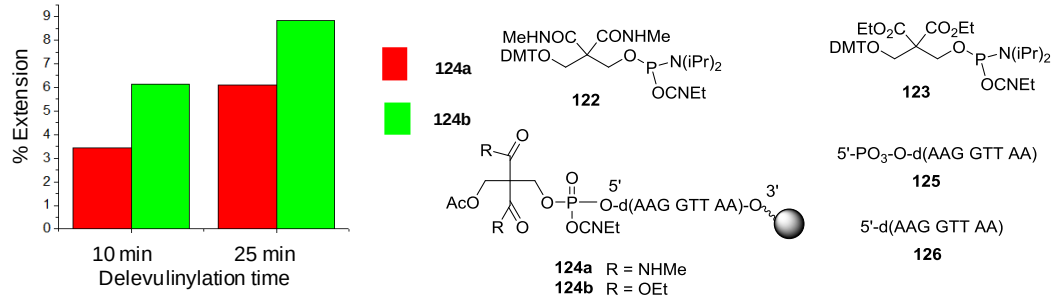
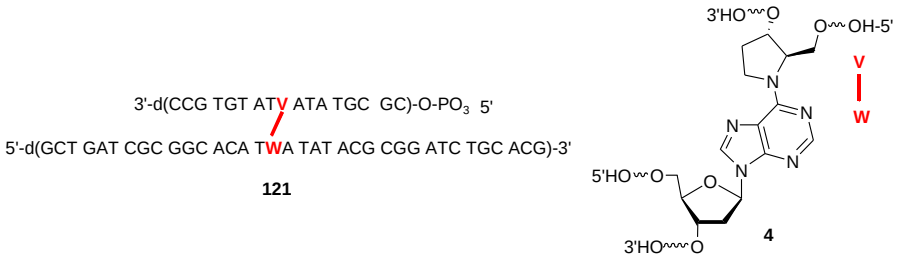


Figure 30. Compatibility of phosphorylating agents **122** and **123** with delevulinylation treatment (0.25 M hydrazine in 3:2 pyridine/acetic acid, room temperature for specified time). Percentage of extension product formed compared to starting material **(124a**,**b)** plotted against delevulinylation time. Only bands corresponding to starting oligonucleotide (**125**) and final extension products were taken into account.

unstable to hydrazinolysis, generating unwanted extension products (branching) (App. Figure 16A and B). The slower moving bands were confirmed as the unwanted extension products using MALDI-TOF MS (App. Figure 16). To assess the exact nature of the effect of hydrazinolysis on the phosphorylating groups, the extension products from 10 and 25 min of hydrazinolysis were quantified with respect to the starting material. It revealed that branching was less for **124a** compared to **124b** for both durations (Figure 30). Furthermore, characterization of bands from the gel by MALDI-TOF MS showed

the presence of unphosphorylated **126** along with phosphorylated **125**, indicating inefficient coupling for **123** (App. Figure 16B and D). Therefore, we chose **122** based on the greater efficiency in solid phase coupling and less unwanted extension (branching).

3.1.2.1.4. Synthesis of longer C4-AP ICL mimic (template strand length ≥ 25 nucleotides)



Next, we strived to optimize the solid phase coupling procedure for minimizing the deletion products similar to those detected during the synthesis of **117**. As described before, stringent strategies were adopted during the synthesis of the template strand, including double coupling of native phosphoramidites and double manual coupling of C4-AP ICL analogue phosphoramidite (**113a**). Anticipating the machine aided automated synthesis of the 5'- and 3'- arms as a possible source of inefficient coupling (thereby generating deletion products), we resorted to manual coupling. During the synthesis of both arms, all phosphoramidites were double-coupled excluding the first, which was coupled four times. The coupling yield was computed based on the dimethoxytrityl cation content after deblocking (detritylation). The amount of resin used, its loading and overall yield during synthesis of the template strand was accounted for within the theoretical yield. Quantification of the trityl cation indicated that the first coupling of the 5'-arm occurred with 87% yield, while the rest of the arm was synthesized with 99.8% average stepwise yield. When the oligonucleotide following 5'-arm synthesis was subjected to ammonolysis, cleavage from support, 5'-^{32}P-labeling and analysis on a

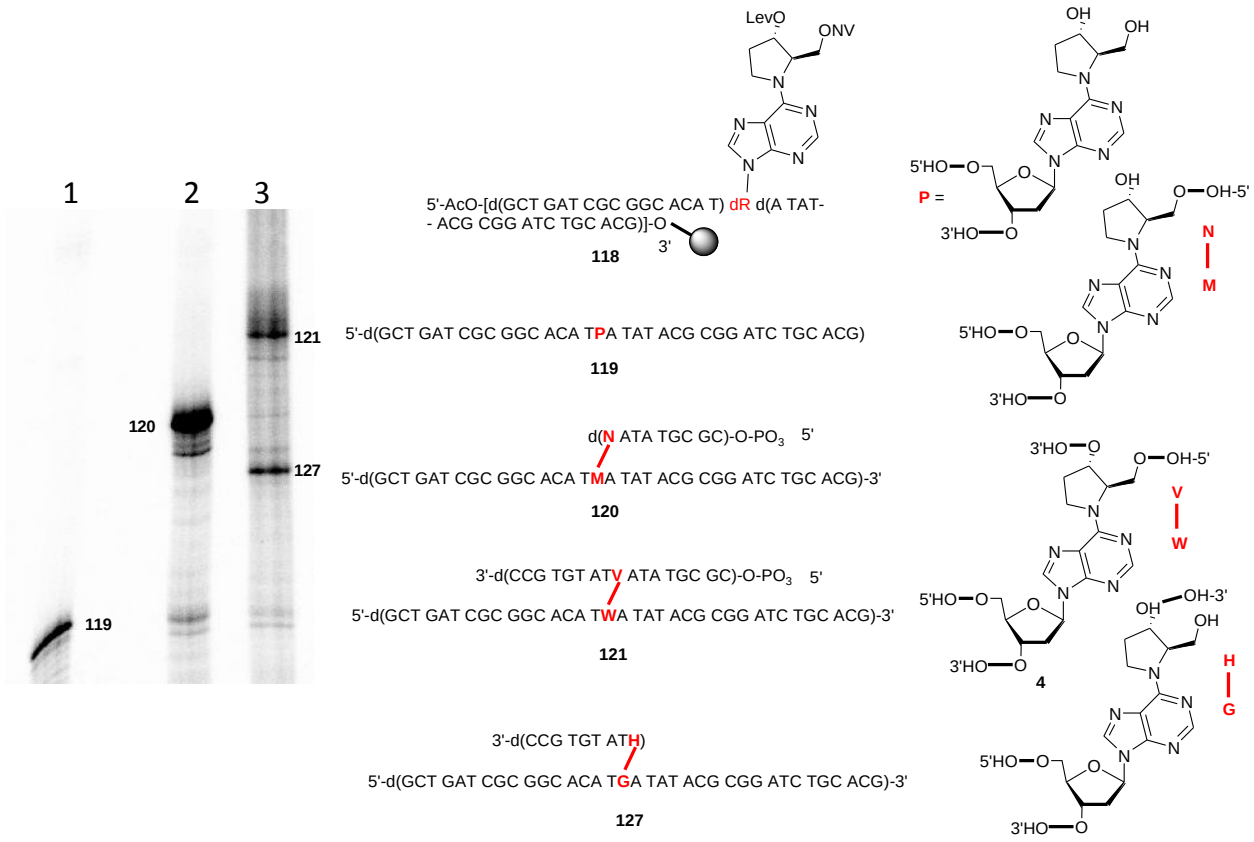


Figure 31. Synthesis of ICL **121** from support bound **118**. Oligonucleotides after synthesis of 5'- and 3'-arms were deprotected and cleaved from the support by ammonolysis. An aliquot from the crude was 5'-^{32}P-labeled, and analyzed on 15% denaturing PAGE gel. Positions of template strand **119** (from unmodified **118**), ICLs **120**, **121**, and **127** are annotated. Lane 1, Template strand (from unmodified **118**). Lane 2, photolysis in toluene (2.5 h, room temperature) and synthesis of 5'-arm using "regular" phosphoramidites. Lane 3, delevulinylation (0.25 M hydrazine in 3:2 pyridine/acetic acid, room temperature for 25 min) and synthesis of 3'-arm using "reverse phosphoramidites.

15% gel, the efficient synthesis of 5'-arm was evident (Lane 2, Figure 31). Similarly, following removal of the 3'-levulinyl group, the 3'-arm was synthesized using manual coupling the same way as the 5'-arm. However, delevulinylation also deprotected some of **118** that was not cleaved upon photolysis, leading to generation of **127** along with desired **121** upon solid phase extension (Lane 3, Figure 31). As a result, quantification of the dimethoxytrityl cation led to coupling yield > 100%. Apart from that, synthesis of **121** was achieved without significant generation of deletion products (Lane 3. Figure 31). ICL **121** was deprotected with concentrated aqueous ammonium hydroxide at 55 °C overnight

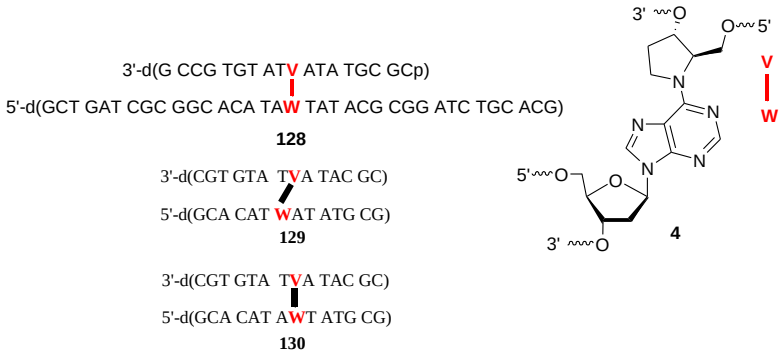
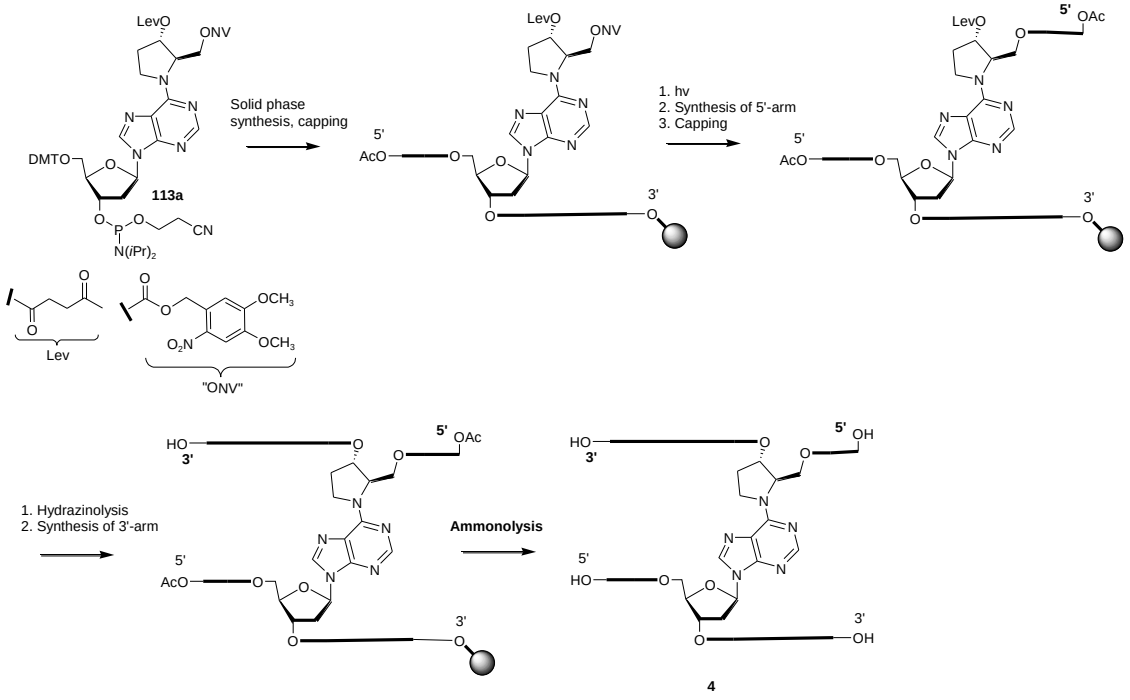


Figure 32. Chemically synthesized cross-linked oligonucleotides containing C4-AP ICL mimic (**4**).

and purified by denaturing PAGE, producing 5.7 nmole of the desired ICL from half of the resin present in a 1 μmol scale column. Using the strategies developed above, ICLs **128** – **130** were synthesized (Figure 32), which were characterized by ESI-MS (App. Figure 114 – 116).

Scheme 35. Summary of C4-AP ICL analogue 4 synthesis using solid phase synthesis

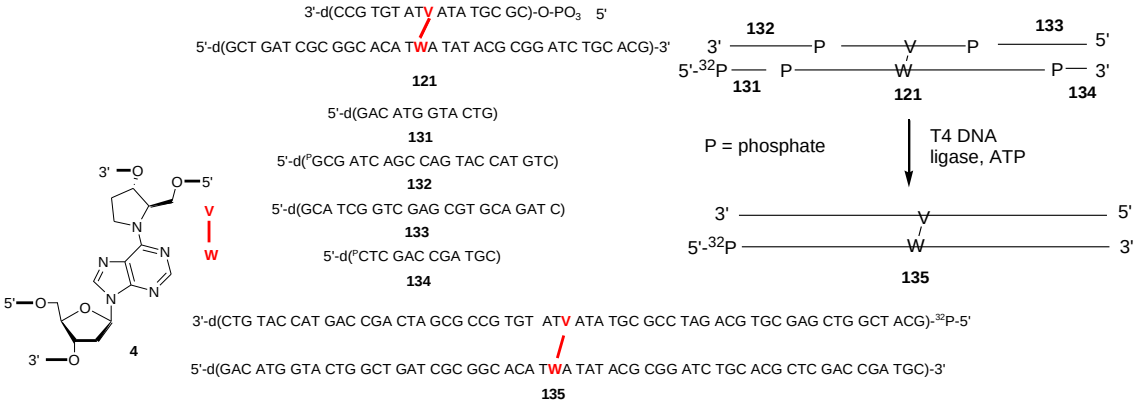


In summary, we established strategies for the C4-AP ICL mimic (**4**) synthesis using the *o*-nitroveratryloxycarbonyl (ONV) and levulinyl (Lev) groups. The synthesis starts with making of the 5'-O-acetyl capped template ("bottom") strand, where the

phosphoramidite **113a** is incorporated via solid phase synthesis (Scheme 35). Next, we determined the viability of removing the ONV group protecting the 5'-hydroxyl of the pyrrolidine, followed by the synthesis of the corresponding arm. After capping the nascent arm, the levulinyl group was removed by hydrazinolysis. Then we employed 5'-cyanoethyl "reverse" phosphoramidites to prepare the 3'-arm, following which ammonolysis yielded the final cross-link product. Manual coupling was employed during synthesis of the arms, while coupling time and counts were optimized to minimize deletion products. The strategy enabled successful synthesis of short (template strand length 8 – 15 nucleotides) and long (template strand length 36 nucleotides) cross-linked oligonucleotides containing **4**.

3.1.2.1.5. Conversion of ICL 121 into 60 bp ICL by T4 DNA ligase mediated ligation

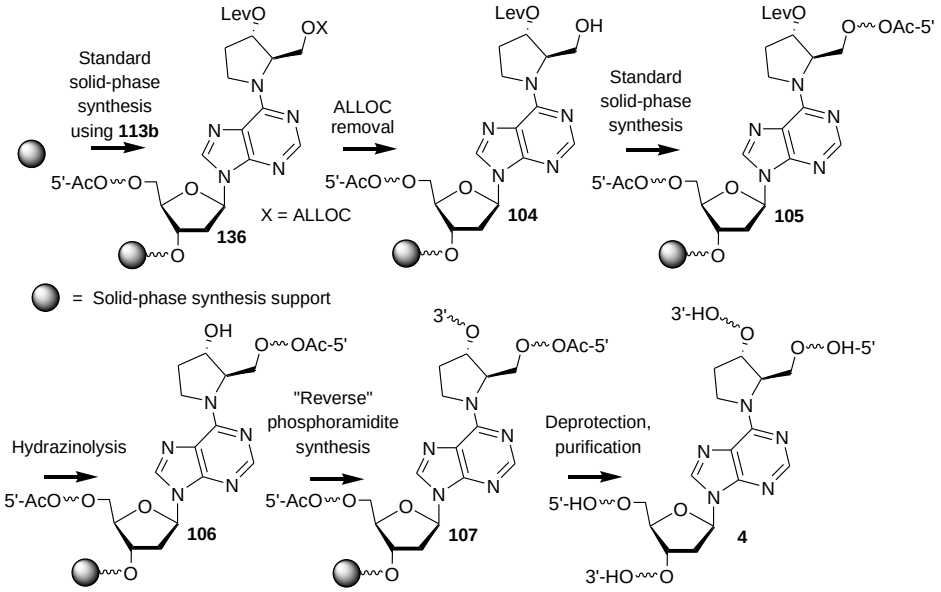
Scheme 36. Synthesis of 60 bp ICL 5'-^{32}P-b-135 by ligation



After synthesizing the C4-AP ICL mimic **121**, we explored the possibility of converting it to full-length cross-link **135** containing fully complementary strands and 5'-^{32}P-labeled bottom strand (cross-linked abasic site containing strand is referred to as "top" strand or "t", while the cross-linked A containing strand is referred to as "bottom" strand or "b") (Scheme 36). Ligation using T4 DNA ligase on **121** and **131-134** enabled

us to synthesize **135**. It required phosphorylation of the 5'-end of **121** that was already not phosphorylated, and was carried out enzymatically using T4 polynucleotide kinase and

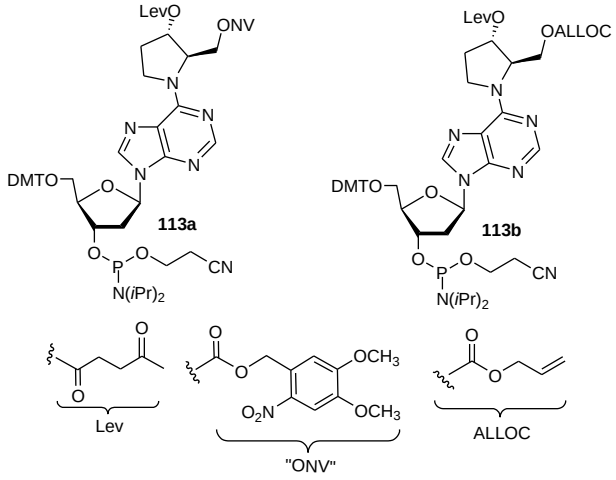
Scheme 37. Synthesis of C4-AP ICL mimic by orthogonal deprotection strategy



ATP (see experimentals for further details). In addition, the 5'-terminus of **133** was independently ^{32}P-labeled with the same enzyme and γ-^{32}P-ATP. Following hybridization of these oligonucleotides, they were ligated (16 °C, 12 h) using T4 DNA ligase and ATP. Oligonucleotide 5'-^{32}P-**133** was used as the limiting agent in the hybridization and ligation (App. Figure 17B). The ligated full length product was purified by 20% denaturing PAGE and isolated from the gel by the crush and soak method. The isolated yield varied between 10 – 20% and was determined based on the specific activity of 5'-^{32}P-**133**. Desalting and rehybridization produced pure 5'-^{32}P-**135**, which was sufficiently long (60 bp) to be used in nucleotide excision repair experiments. The integrity of 5'-^{32}P-**135** was confirmed by using restriction enzymes (App. Figure 17A and C). The ligation mediated construction can be employed to selectively label other termini of **135** as well.

3.1.2.2. Synthesis of C4-AP ICL mimic from 5'-ALLOC-3'-Lev-C4-AP phosphoramidite

While exploring conditions to synthesize C4-AP ICL mimic using 5'-*o*-nitroveratryloxycarbonyl-3'-levulinyl-C4-AP phosphoramidite (**113a**), we investigated the opportunities of 5'-allyloxycarbonyl-3'-levulinyl-C4-AP phosphoramidite (**113b**) for



the same. As explained before, we functionalized the 3'-pyrrolidinyl OH using levulinyl (Lev) due to its better deprotection efficiency and inefficient 3' to 5' coupling. Similar to *o*-nitroveratryloxycarbonyl (ONV) mediated synthesis of C4-AP, the beginning of the

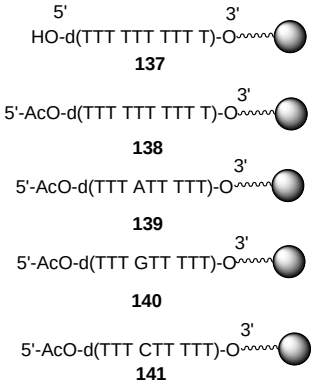


Figure 33. Oligonucleotides used to determine compatibility of *O*-acyl and exocyclic amine protection with Pd(0) catalyzed allyloxycarbonyl (ALLOC) removal.

synthesis proceeded by incorporation of the requisite phosphoramidite (**113b**) into template strand by solid phase synthesis. Following removal of the allyloxycarbonyl protection from **136**, the rest of the synthesis leading to **4** was anticipated to proceed via the same strategy as the ONV protected analogue (Scheme 37).

Previously, the allyloxycarbonyl (ALLOC) group protecting nucleotide exocyclic amines was removed using various reagents based on Pd(0) chemistry.[138] To the best of our knowledge, an example of ALLOC deprotection followed by solid phase synthesis on oligonucleotide substrates has never been reported. Similar chemistry was reported during solid phase peptide synthesis, which we adapted for our cause.[139,140] The acyl capping and exocyclic amine protecting groups needed to withstand the Pd(0) deprotection conditions, while the subsequent solid phase extension should be unaffected by it. To test their robustness towards Pd(0) based deprotection conditions, we subjected **136** to ALLOC removal ($Pd(PPh_3)_4$ (0.1 μM) and $Et_2NH{\bullet}BH_3$ (7 μM) in DCM) and subsequent washing step (using 0.2% trifluoroacetic acid in DCM and 5% DIPEA in DCM, see experimental for details). Following Pd(0) treatment and washing, solid phase synthesis using 5'-dimethoxytrityl "regular" phosphoramidites was applied to **136**. If the Pd(0) treatment and washing is compatible with solid phase synthesis, it would lead to efficient extension of **136**. The efficiency or failure of the extension was probed by ammonolysis of the oligonucleotide on the support, ^{32}P-labeling, and analysis on 20% denaturing PAGE. The analysis revealed inefficient extension compared to that without Pd(0) treatment (App. Figure 18). Separate treatment of the resin with the Pd(0) solution (as $Pd(PPh_3)_4$) or washing conditions identified the former as the cause of the inefficient solid phase extension (Lane 3 and 4, App. Figure 18). The Pd(0) treatment contained two

components added together, 7 μM $Et_2NH•BH_3$ and 0.1 μM $Pd(PPh_3)_4$ in 1 mL DCM. To pinpoint the possible source of the inefficiency, **136** was independently treated with these components as well as with the washing conditions. Furthermore, the washing conditions adopted from the solid phase peptide synthesis were followed by subsequent washing steps using acetonitrile and methanol (App. Figure 19). This time, however, the solid phase extension following these treatment proceeded efficiently, possibly benefitting from the modified extensive washing steps.

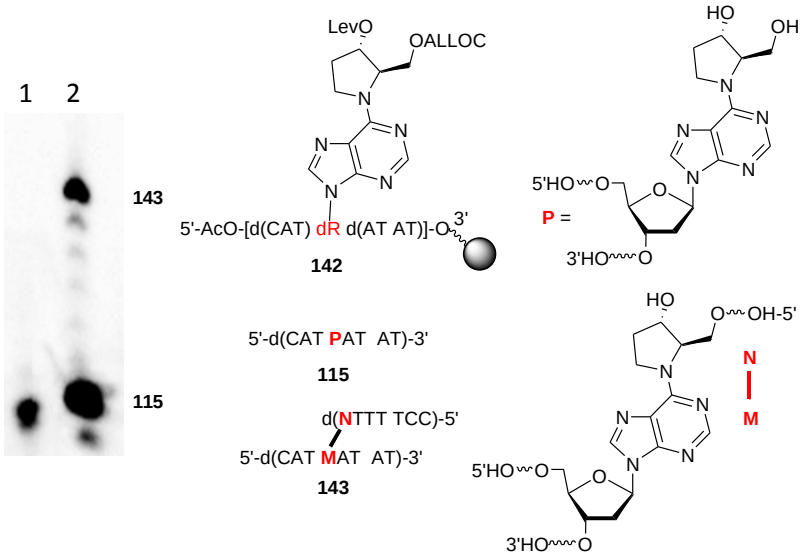


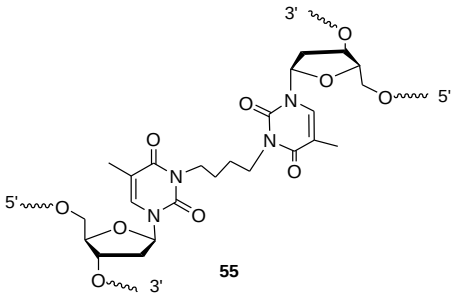
Figure 34. Solid phase oligonucleotide synthesis following ALLOC removal. (A) Oligonucleotides used to determine compatibility of *O*-acyl and exocyclic amine protection with Pd(0) catalyzed allyloxycarbonyl (ALLOC) removal. (B) Synthesis of ICL **143** from support bound **142**. Following synthesis, oligonucleotides were deprotected and cleaved from the support by ammonolysis. An aliquot from the crude was 5'-^{32}P-labeled, and loaded on 20% denaturing PAGE gel. Positions of template strand **115** (from unmodified **142**), and ICL **143** are annotated. Lane 1, Unmodified **136** deprotected. Lane 2, **142** subjected to Pd(0) catalyzed ALLOC removal, followed by solid phase extension using "regular" phosphoramidites.

Having verified the compatibility of Pd(0) deprotection and subsequent washing steps with solid phase synthesis, the stability of acetyl capping under the same conditions was explored using **137**. After Pd(0) treatment, washing, subsequent solid phase extension, and ammonolysis of the resin, an aliquot from the crude was probed by ^{32}P-labeling and denaturing PAGE to detect any unwanted extension from the removal of

acetyl (App. Figure 20). No extension product was detected implying stability of acetyl under this condition. Next, we sought to probe the compatibility of exocyclic amine protections using **138** – **140**. Probing the extension using ^{32}P-labeling and subsequent denaturing PAGE analysis demonstrated that the amine protections (benzoyl-A, isobutyryl-G and benzoyl-C) were stable to the ALLOC deprotection conditions (App. Figure 21).

Having established the compatibility of the ALLOC deprotection conditions with acetyl and phosphoramidite amine protections, as well as the solid phase synthesis conditions, we attempted to synthesize the 5'-arm using this chemistry. When Pd(0) deprotection of **142** and subsequent solid phase synthesis to produce **143** was probed by ^{32}P-labeling and denaturing PAGE, 37.5% extension was detected (Figure 34). The identity of ICL **143** was confirmed by MALDI-TOF MS (App. Figure 117). Although the synthesis revealed significant presence of template strand **115** from unmodified **142**, the purity of **143** as observed on the gel and by MALDI-TOF MS was satisfactory. We could not account for the incomplete extension despite the previously established compatibility of solid phase synthesis with Pd(0) based ALLOC removal. In our opinion, the 20 min Pd(0) treatment time as described in solid phase peptide synthesis methodology could be too short for ALLOC removal from an oligonucleotide. Before experimenting with the treatment duration, we decided to establish whether the isolated **143** had the correct positioning of the cross-link. We investigated this by subjecting 3'-^{32}P-**143** to hydroxyl radical cleavage treatment (App. Figure 22). While the treatment revealed correct connectivity of the cross-link (on the fifth nucleotide from the 3'-terminus), it showed the presence of multiple bands for untreated **143**, indicating decomposition. Having

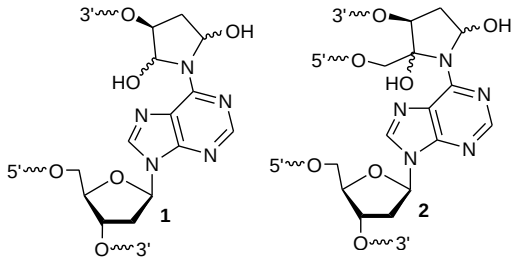
previously established the purity of **143** by MALDI-TOF MS, the decomposition was paradoxical in nature. By this time, our parallel endeavors with 5'-*o*-nitroveratryloxycarbonyl protection successfully delivered ICL analogues containing **4**. Therefore, the inefficient extension observed for 5'-ALLOC, combined with unexplained decomposition of the isolated material led us to abandon any further attempts using this route.



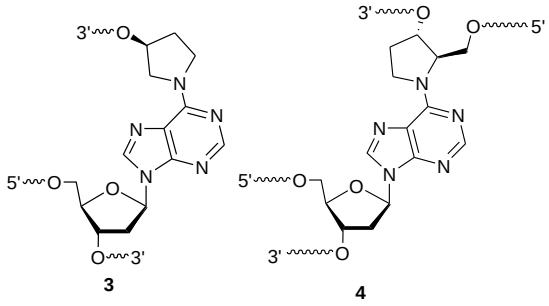
55

After we abandoned attempts of making **4** using orthogonal deprotection of ALLOC, we came across the synthesis of *N3*-thymidine-butylene-*N3*-thymidine ICL (**55**).[33] The synthesis of **55** was reported prior to our attempts and we were unaware of this article, which successfully utilized the orthogonal ALLOC removal.[33] In contrast to our method, the cleavage of 5'-ALLOC employed treating the resin with $Pd(PPh_3)_4$ (3 mM), PPh_3 (5.8 mM) in 1.2 M n-butylamine/formic acid buffer in 1 mL THF (35 °C, 3 h).[33] The removal of the Pd(0) from the resin was carried out by washing it with saturated N,N-diethyldithiocarbamate, which acted as a chelating agent for palladium.[33] Thus, the employment of higher concentration of Pd(0), as well as elevated temperature, and longer time (3 h versus 20 min) in combination with superior washing method could have led to better solid phase synthesis efficiency than ours.

3.1.3. Summary

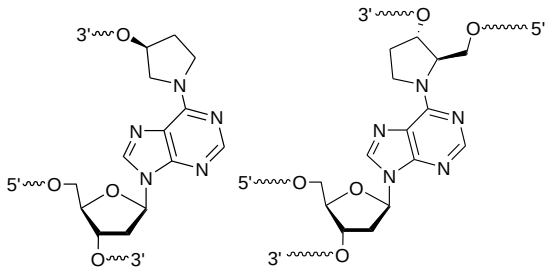


The synthetic strategies discussed enabled successful synthesis of stable analogues of DOB and C4-AP ICLs (**1** and **2**). The native cross-links (**1** and **2**) contain hemiaminal linkages, leading to their instability.[9,31] The stable analogues (**3** and **4**) are the reduced version of the native ICLs, missing the hemiaminal linkages. While the native ICLs contain epimerizable centers, the stereochemistry of the native deoxyribose sugars



are retained in the stabilized cross-link analogues, assuming they would be the least distorting. We demonstrated the compatibility of a pair of protecting groups (*o*-nitroveratryloxycarbonyl and levulinyl) with solid phase oligonucleotide synthesis, and implemented them to synthesize **3** and **4**. In combination with enzymatic ligation, this strategy enables synthesis of cross-linked oligonucleotides long enough (e.g. 60 bp) for nucleotide excision repair studies. In addition, the strategies have general applicability for synthesizing cross-linked or branched DNA and RNA.

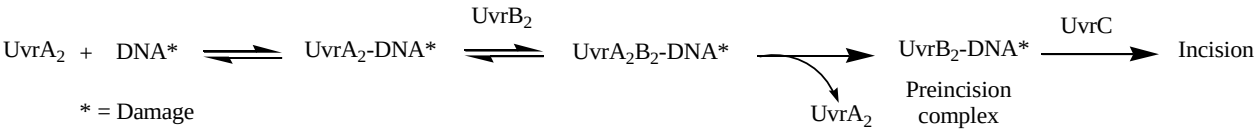
3.2. Nucleotide excision repair studies on DOB and C4-AP ICL



3 **4**

Among UvrA, UvrB and UvrC, the bacterial nucleotide excision repair proteins, the first two act as a damage sensor, while UvrC incises an oligonucleotide containing the lesion. The preliminary damage recognition is done by UvrA, which binds damaged DNA as a dimer (Scheme 38). After an initial and not so thorough probing by $UvrA_2$, $UvrB_2$ joins the $UvrA_2$-DNA complex in the ongoing damage search and performs a final

Scheme 38. Binding and incision of damaged substrate by UvrABC



and confirmatory scan. After finalizing that the DNA in question is genuinely damaged, $UvrA_2$ departs from the search machinery (Scheme 38). At the same time, $UvrB_2$ forms a stable complex with DNA (the preincision complex). The formation of preincision complex is regarded as the rate determining step of UvrABC incision.[124] In the next step, UvrC binds the preincision complex and cleaves the DNA. In our study, we probed the interaction of UvrABC with DOB and C4-AP ICL analogue (**3** and **4**). We also investigated the efficiency of preincision complex formation for both lesions and attempted to correlate that with incision yield.

3.2.1. Nucleotide excision repair substrate designs and preparations

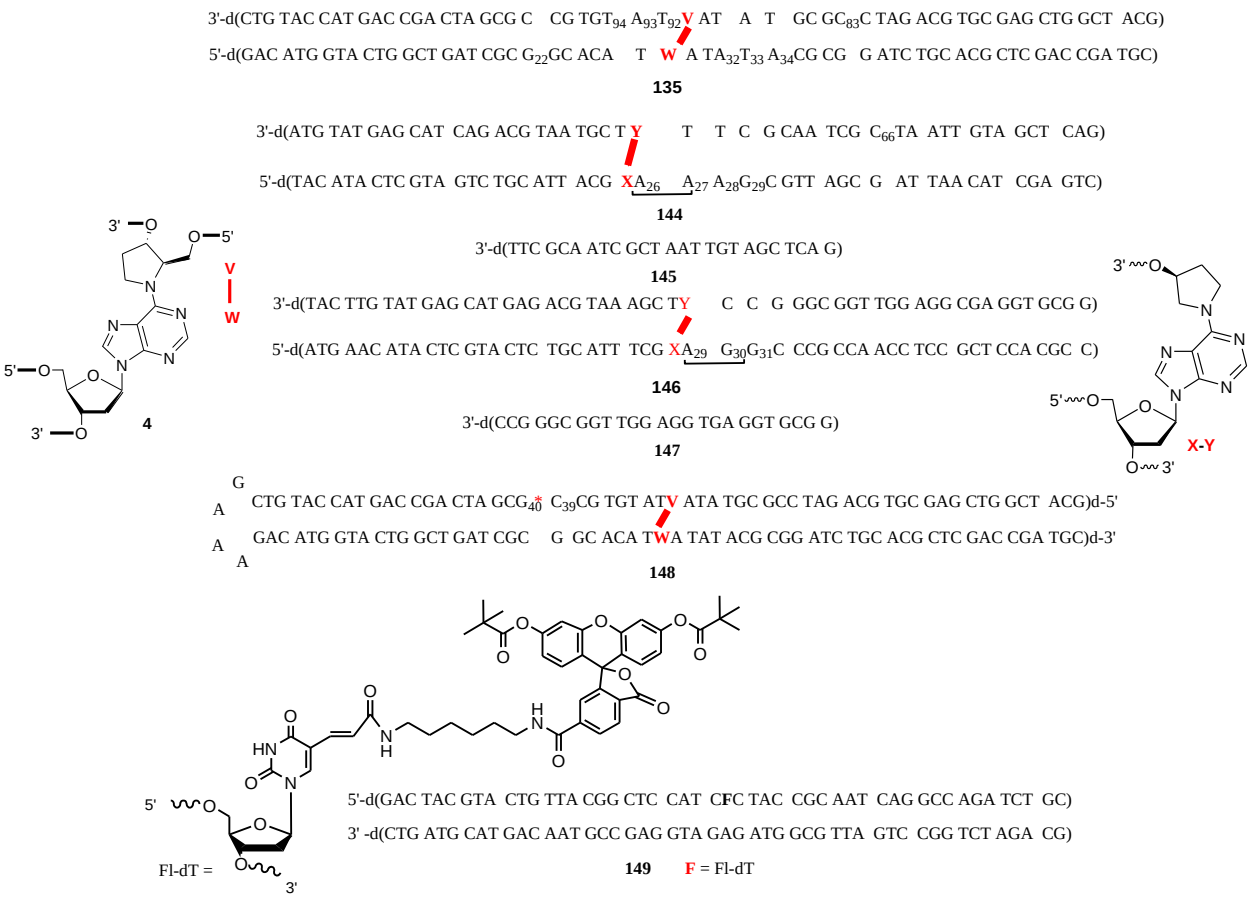
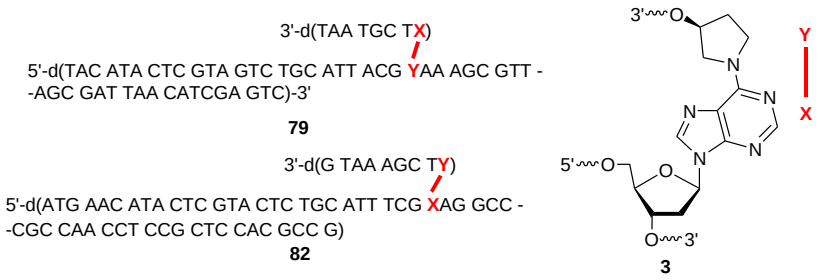


Figure 35. Oligonucleotides used in nucleotide excision repair studies.

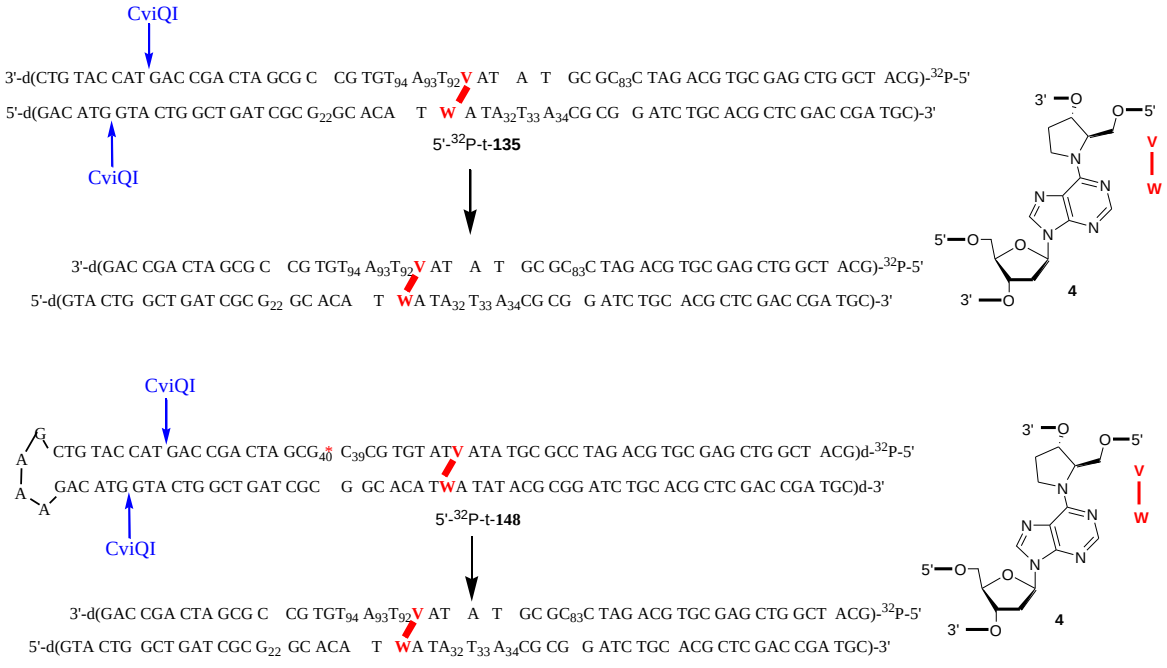
The lengths (>50 bp) of C4-AP and DOB ICL mimics used for repair studies were adequate for UvrABC binding and incision as determined from the footprinting of $UvrA_2B_2$-DNA complex formation.[111] The construction of C4-AP ICL substrate **135** was independently ^{32}P-labeled at all four termini by enzymatic ligation (Scheme 36 and App. Figure 17, see experimentals for more details). We prepared cross-linked duplex **144** containing DOB ICL analogue **3** from precursor **79** by enzymatic ligation (see experimental for further details). After conducting incision studies, a standard T_m calculator (www.idtdna.com/analyzer/Applications/Oligoanalyzer/) suggested the instability of the un-cross-linked region of **144** at NER reaction conditions (2 nM

substrate concentration, 55 °C). This was further tested by incubating 2 nM 5'-^{32}P-t-**144** (the strand containing abasic site analogue in each substrate is labeled "t" or "top" while that containing the cross-linked A is denoted as "b" or "bottom") with unlabeled **145** (10 nM) at 55 °C without UvrABC. Analysis of aliquots removed over the course of 60 min



in nondenaturing PAGE (20%) revealed up to 28% melting (detected by accumulation of 5'-^{32}P-**145**) at 60 min, corroborating that **144** was unstable under the UvrABC reaction conditions (App. Figure 23). We therefore prepared **146** with greater G-C content in the

Scheme 39. CviQI cleavage of 5'-^{32}P-t-148



un-cross-linked duplex region to prevent unwanted denaturation. Similar to **144**, **146** was prepared and separately labeled at all four termini by enzymatic ligation from its precursor **82** (see experimental for further details). The integrity of the full length cross-linked **146** was confirmed by restriction enzyme digestion (App. Figure 25). An estimation using the standard T_m calculator indicated the stability of its un-cross-linked region (T_m ~75 °C) under UvrABC reaction conditions. Its hybridization persistence was

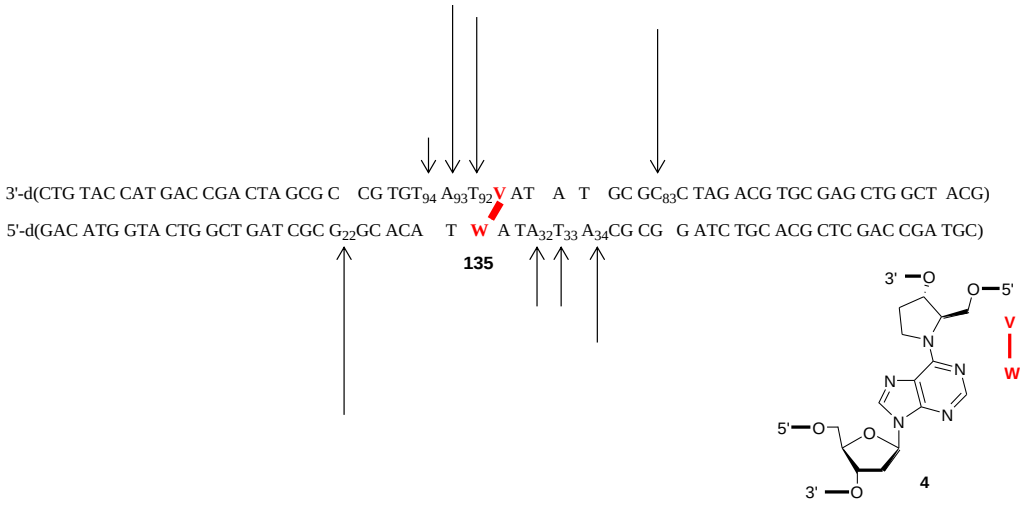


Figure 36. Histogram describing UvrABC incision of C4-AP ICL analogueue **4** in **135**. Data taken from four individual experiments in which each terminus of **135** was ^{32}P-labeled.

further verified by incubation of 5'-^{32}P-t-**146** (2 nM) with unlabeled **147** (10 nM). Aliquots were removed over the course of 60 min and analyzed by 20% nondenaturing PAGE. There was no accumulation of 5'-^{32}P-**146** or loss of 5'-^{32}P-t-**146** at 60 min, confirming that **146** was stable at UvrABC reaction conditions (App. Figure 26). The incisions of **143**, **144**, and **146** were compared with that of a duplex containing a C5-fluoresceinylated thymidine derivative (**149**), a standard substrate for comparing enzyme activity. In addition, substrate **148** containing ICL analogue **4** was prepared with a hairpin, and tagged with a ^{32}P-label at the 5'-end of the top strand (5'-^{32}P-t-**148**) or internally at the phosphate between C_{39} and G_{40}. Migration of CViQI restriction enzyme

digest of 5'-^{32}P-t-**148** and 5'-^{32}P-t-**135** were compared on denaturing PAGE to confirm the integrity of **148**, since both would generate the same radiolabeled product (Schemes 39, App. Figures 27 and 28). Furthermore, treatment of internally labeled **148** with restriction enzymes Fnu4HI (producing a 44mer product) and HpyCH4V verified its full-length structure and hybridization (App. Figure 29).

3.2.2. UvrABC incision of C4-AP ICL analog (4)

UvrABC incision sites on C4-AP ICL analogue were determined based on the incubation of the protein in four separate experiments where **135** was independently ^{32}P-

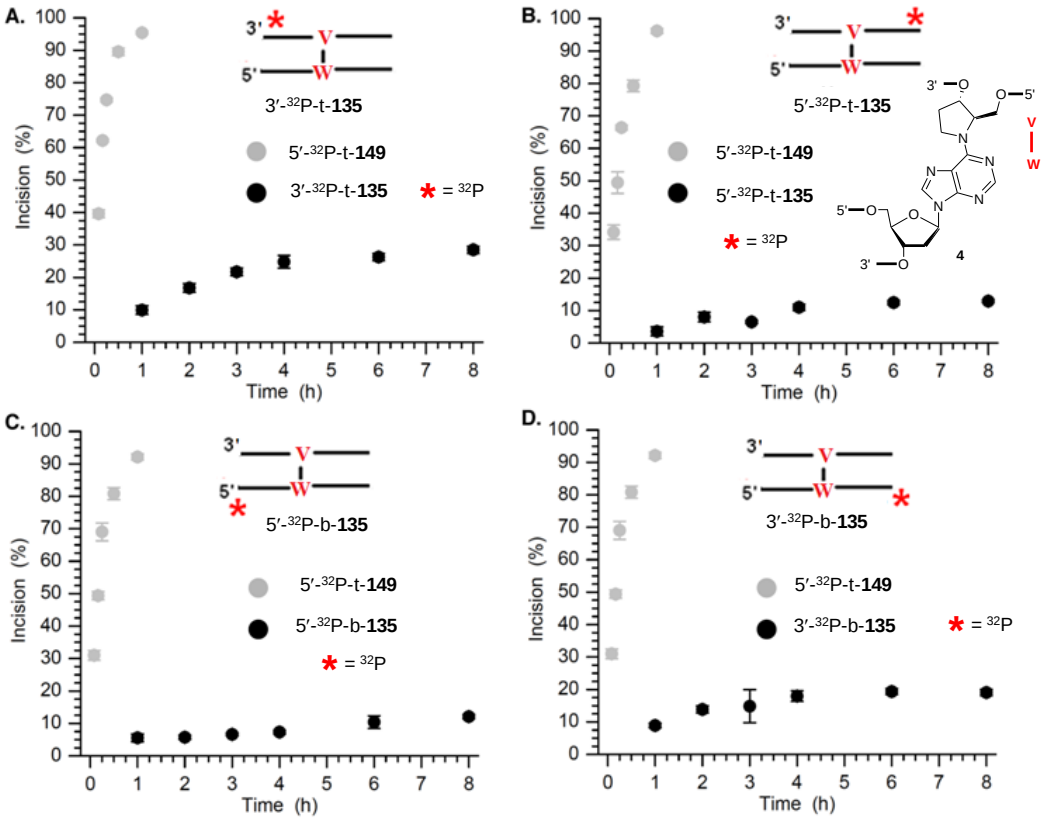
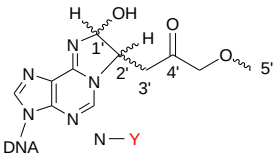


Figure 37. Time dependence of UvrABC incision of C4-AP ICL analogue **4** in ^{32}P-labeled **135** compared to that in 5′-^{32}P-**149**. The total amount of incision at all nucleotides in the region between the ^{32}P label and the cross-link is plotted: (A) 3′-^{32}P-t-**135**, (B) 5′-^{32}P-t-**135**, (C) 5′-^{32}P-b-**135**, and (D) 3′-^{32}P-b-**135**.[141]

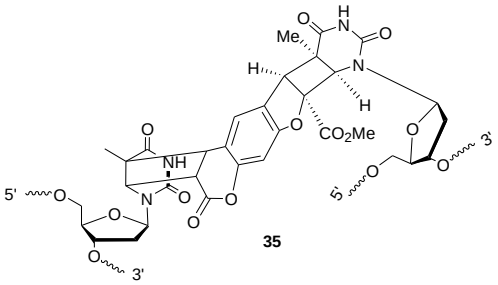
labeled at each terminus. For the strand containing the abasic site analog ("top" strand), only one nucleotide was incised on 5'-side of the lesion (C_{83}), while three nucleotides were incised on the 3'-side of it (T_{92}, A_{93} and T_{94}, Figure 36 and App. Figure 30). It should be noted that a 5'-labeled substrate is incised to produce a 3'-OH causing it to migrate slower than the equivalent residue in the sequencing lane, which contain a 3'-phosphate terminus (App. Figure 30).[128] For the strand containing the cross-linked A analogue ("bottom" strand), incisions at lone G_{22} on the 5'-side, and A_{32}, T_{33} and A_{34} on the 3'-side of the lesion were observed (Figure 36 and App. Figure 31). The quantification of the yield was done on the basis of incision between the cross-link and the ^{32}P label, although bands from other incisions could be visualized on the gel. For the "top" and "bottom" strand, the 5'-incision occurred eight and seven nucleotides from the cross-linked sites with cleavage yields of 16% and 13.5% at 8 h, respectively. In contrast, the 3'-incision on the "top" and "bottom" strands of **135** occurred at three positions at each strand and was considerably closer to the respective cross-linked positions. For the "top" strand, the incisions on the second and third nucleotides (T_{92} and A_{93}) starting from the cross-linked abasic site analogue accounted for the majority of the 27-28% cleavage observed after 8 h (Figure 36). At the same time, the 16-18% cleavage observed after 8 h on the 3'-side of the cross-linked A in the "bottom" strand was more evenly distributed over A_{32}, T_{33} and A_{34}, while the incision sites shifted further from the cross-link site. Despite the shift, the size of the unhooked oligonucleotide patch was similar for incisions on the top and bottom strands of **135**, 10 – 12 nucleotides and 11 – 13 nucleotides respectively (Figure 36).

The incision yield of a typical substrate by UvrABC is usually compared with that of the standard substrate, a duplex containing fluoresceinylated T (**149**). The comparison demonstrated that the incision for **135** was significantly less compared to that for **149** (Figure 37). While greater than 90% of 5'-^{32}P-t-**149** was cleaved by 60 min, 27-28% incision was observed for 3'-^{32}P-t-**135** in 8 h, and less so for other regions of **135** (Figure 37). The cleavages on the 3'-side of the cross-linked nucleotides were greater than that on the 5'-side, with the incision on the 3'-side of top strand being almost twice that of 5'-sides.



13

C4-AP thermodynamic ICL (**13**) in DNA



35

UvrABC incisions on an ICL typically occur at the third phosphodiester bond 3' to the ICL and ninth phosphodiester bond 5' to the ICL, producing a 14 nucleotide long segment.[120] Although the length of the nucleotide patch resembled that produced by incision of the C4-AP ICL, the sites of incision differed slightly. Instead, the 5'-incisions were at the eighth and seventh phosphodiester bonds in the top and bottom strands, respectively. At the same time, the 3'-incisions were distributed between the first and third phosphodiester bonds on the top strand and third and fifth phosphodiester bonds on the bottom strand, relative to the cross-linked nucleotide. The nature of the incisions was asymmetric, a fact that was previously encountered during UvrABC incision of the C4-AP thermodynamic cross-link (**13**)[22] and psoralen cross-link (**35**).[128] During the UvrABC incision of psoralen cross-link, 70-85% G-C content on 5' side and 50% G-C content on

3' side of the furan adducted strand led to predominant or exclusive incision on that side.[128] Despite the fact that the top strand of **135** had a greater overall incision yield, less than 60% G-C content in both its 5' and 3'-side ruled out any possible correlation between G-C content and strand preference.

3'-d(CTG TAC CAT GAC CGA CTA GCG CCG TGT AT)
5'-d(GAC ATG GTA CTG GCT GAT CGC G)
150

5'-d(GCA TCG GTC GAG CGT GCA GAT C)
3'-d(CGT AGC CAG CTC GCA CGT CTA GGC GCA TA)
151

5'-d(GCA TCG GTC GAG CGT GCA GAT)
3'-d(CGT AGC CAG CTC GCA CGT CTA GTG TTT TCC GAT CAG ATC GAA)
152

5'-d(AAG CTA GAC TAG CCT TTT GTG ATC TGC AC CTC GAC CGA TGC)
153

5'-d(TAT GTC CCG CGA TCA GCC AGT ACC ATG TCA AAG GAC ATG GTA CTG GCT GAT CGC G)
154

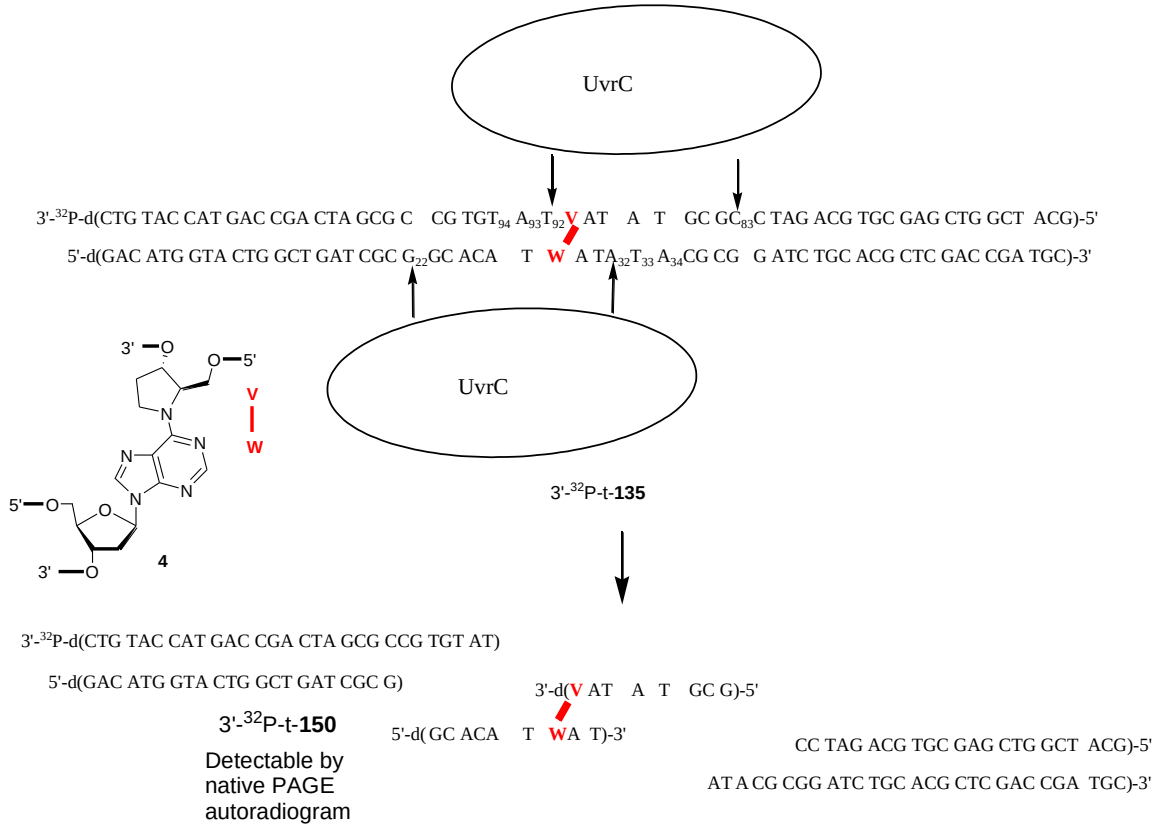
Figure 38. Oligonucleotides used for detection of double strand break during UvrABC incision of C4-AP ICL analogue **4**.

3.2.3. Detection of double strand break during UvrABC incision of C4-AP ICL analog (4)

UvrABC incision on A-T and mitomycin C ICLs reportedly produces double strand breaks (DSBs) by repetitive cleavage.[20,21] Since **135** was incised on both strands by UvrABC, we determined whether this generates a DSB, or whether reaction occurs on opposite strands in separate cross-linked molecules. UvrABC incision of 3'-^{32}P-t-**135** was loaded on a 20% native PAGE, along with the anticipated DSB product **150** labeled at its top strand (3'-^{32}P-t-**150**, Scheme 40, Figure 38) in a separate lane. Detection of a band that migrated similarly to 3'-^{32}P-t-**150** indicated 1 – 3% DSB formation (Figure 39). This also implied a single to double strand break ratio of ~10:1, which is slightly greater than what was observed during the repair of the C4-AP thermodynamic cross-link (**13**).[22] We

wondered whether this product is a real product from DSB or an artifact resulting from the renaturation of products generated from two independent single-strand incisions. To

Scheme 40. Repetitive incision on 3'-^{32}P-t-135 to produce double strand break



determine the authenticity of the band corresponding to the DSB, we designed substrate **151** (Figure 38), a potential DSB product from incisions on the 5'-side of the top strand and 3'-side of the bottom strand. Duplex **151** was ^{32}P-labeled at the 5'-terminus of the top strand to produce 5'-^{32}P-t-**151**. To monitor its denaturation and potential renaturation with an identical complementary substrate, 5'-^{32}P-t-**151** was incubated separately with 1 and 5 equiv of **153** under the UvrABC reaction conditions (55 °C, 3 h, Figure 39). Oligonucleotide **153** consisted of a complementary region similar to the unlabeled strand of 5'-^{32}P-t-**151**, except it was 13 nucleotide longer. The concentration (0.26 nM) of 5'-^{32}P-

t-**151** was based on the observed incision yield detected for UvrABC treatment of **135**. Upon analysis of the aliquots from the incubation on 16% nondenaturing PAGE, 9.8%

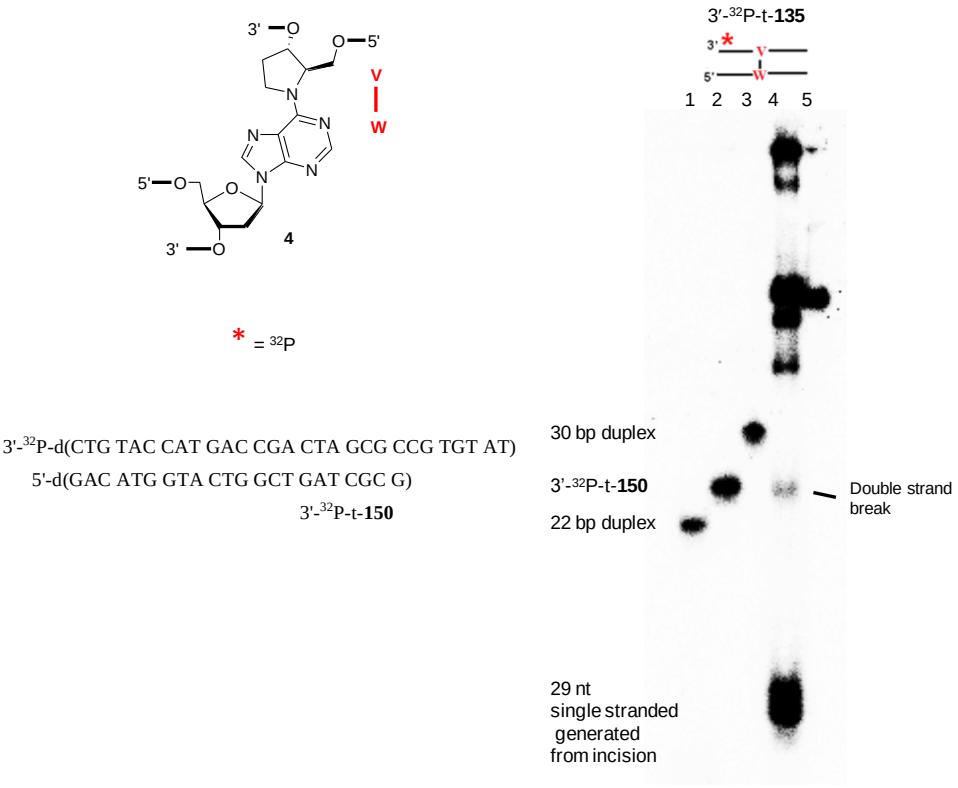
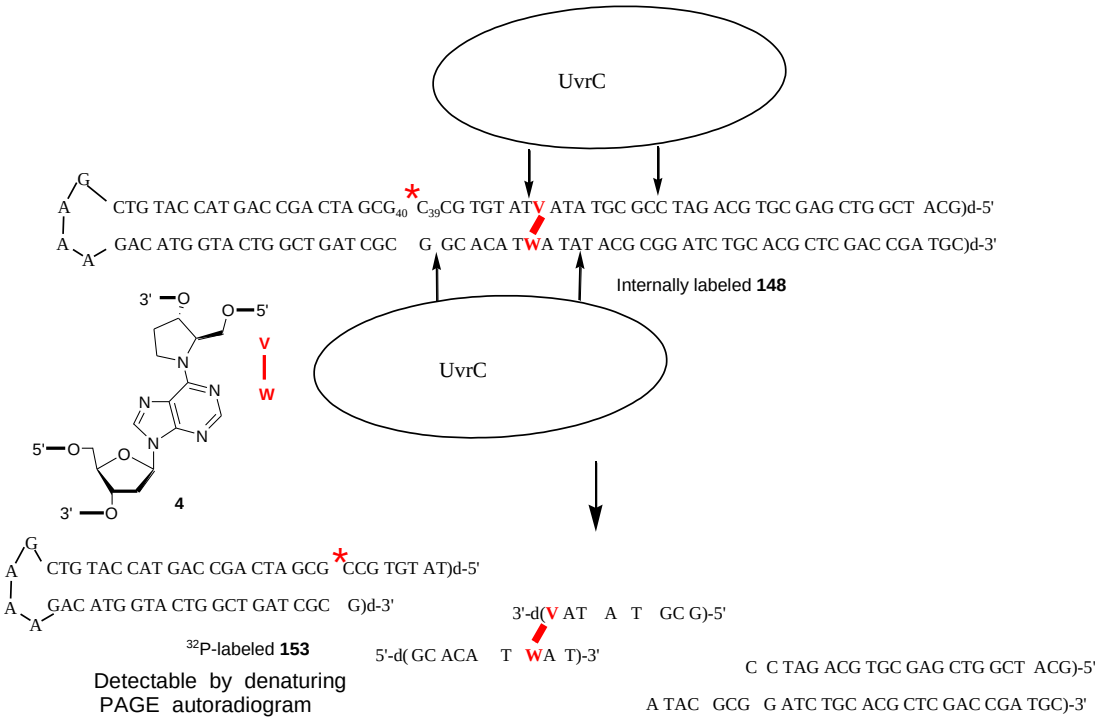


Figure 39. Representative gel showing migration of products from UvrABC incision on 3'-^{32}P-t-**135** in a 20% non-detaturing PAGE. Lane 1, independently synthesized 44mer duplex. Lane 2, oligonucleotide 13 labeled at 5'-side of shorter strand. Lane 3, independently synthesized 60mer duplex. Lane 4, aliquot from UvrABC proteins incubated with 3'-32P-t-**135** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 8 h. Lane 5, 3'-32P-t-**135** without any treatment.

and 21.9% slower migrating exchange products (5'-^{32}P-t-**152**, Figure 38) were observed in the presence of **153** (1 and 5 equivalents, respectively) (App. Figure 32). The exchange experiment clearly indicated that visualization of DSB product from **135** could be from single strand incision reactions on separate substrate molecules. The hairpin substrate **148** was therefore employed to resolve this ambiguity (Scheme 41). It was ^{32}P-labeled at the phosphate between C_{39} and G_{40} or at the 5'-terminus of the top strand (5'-^{32}P-t-**148**). Incubation of UvrABC with 5'-^{32}P-t-**148** cleaved the substrate at the same position and yield as 5'-^{32}P-t-**135** (Figure 40A). The identical location and yield of the incision

confirmed that the presence of the hairpin did not affect its recognition of **4** by the repair proteins. Next, the internally labeled hairpin substrate **148** was treated with UvrABC. We anticipated that in the event of concurrent incisions on the top and bottom strands on the side of the hairpin loop, internally labeled **148** would produce a single 54-nucleotide ^{32}P-

Scheme 41. Repetitive incision on internally labeled 148



labeled product (Scheme 41). Treatment of internally labeled **148** however did not produce any 54-nucleotide long oligonucleotide that migrated like the independently synthesized DSB product 5'-^{32}P-**154** (Scheme 41 and Figure 40B). Instead, both incisions on 5'-^{32}P-t-**148** and internally labeled **148** produced bands that migrated slower than **148** itself (Figure 40). These bands could be from single incision at regions between **4** and the hairpin of **148**, since their migration patterns resembled those from an incomplete ligation of **148** (App. Figure 33).

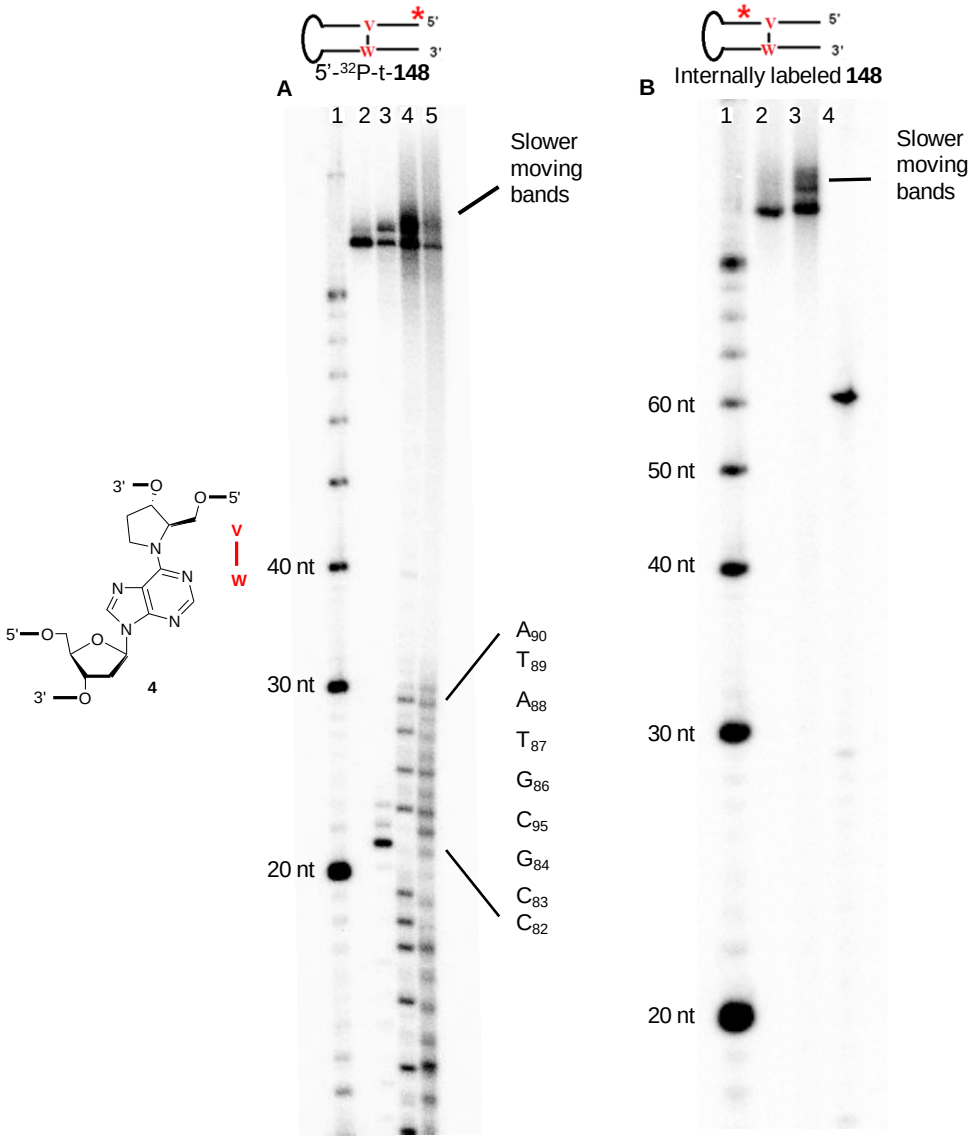
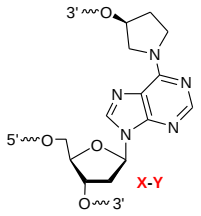


Figure 40. UvrABC incision of hairpin DNA (**148**) containing C4-AP ICL analogueue **4**. (A) Incision of 5′-^{32}P-**148**: lane 1, nucleotide size markers; lane 2, unreacted 5′-^{32}P-**148**; lane 3, incised 5′-^{32}P-**148**; lane 4, A + G reaction; lane 5, T + C reaction. (B) Incision of internally labeled ^{32}P-**148**: lane 1, nucleotide size markers; lane 2, unreacted 11; lane 3, incised 11; lane 4, 5'-^{32}P-**153**. Position of bands migrating slower than **148** are annotated.

3.2.4. Incision of DOB ICL analogue (3) by UvrABC

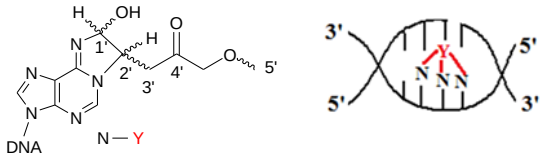
DOB ICL is structurally similar to the C4-AP thermodynamic ICL (**13**), having a strand break adjacent to the cross-link. A misplaced incision (e.g. incision at cross-linked A containing strand) would yield a double strand break (DSB). Predominant incision at

the 3'-side of A ("bottom strand") was detected, which indeed generated double strand breaks in **144** (Figure 41, 42 and App. Figure 34 and 35). In addition, minor cleavage at



3

the 5'-side of the abasic site analogue ("top strand") of the un-cross-linked region was observed (Figure 41, 42 and App. Figure 36). The sites of incision on the bottom strand (A_{26} – A_{28}) were within one to three nucleotides from the cross-linked site, while on the



13

top strand it (C_{66}) was 10 nucleotides away (App. Figure 34 and 36). The amount of incision on A_{27} was considerably higher than that on A_{28}. The overall incision yield at the bottom strand was ~33% after one hour, while the standard substrate fluoresceinylated T

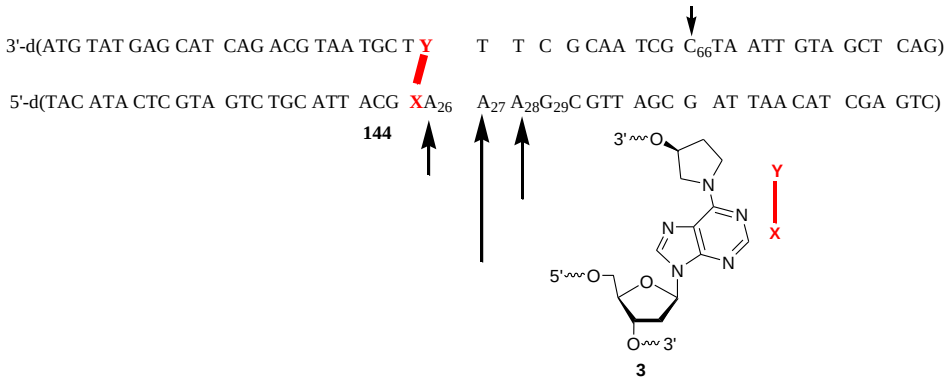


Figure 41. Histogram describing UvrABC incision of DOB ICL analogue **3** in **144**. Data taken from four individual experiments in which one of the four termini of **144** was ^{32}P-labeled.

was incised with a yield between 81 and 90% by that time (Figure 41A and App. Figure 35). In contrast, C_{66} from the "top strand" was incised with a modest ~3% yield in 1 h (Figure 41 and App. Figure 36). We speculated that cleavage at C_{66} was possibly due to the immediate vicinity of a strand break (10 nucleotides away from C_{66}), that influenced the protein to incise a signature 10 nucleotide long fragment. However, we were unable to provide a rationale behind the predominant incisions seen at the 3'-side of the bottom strand leading to DSB.

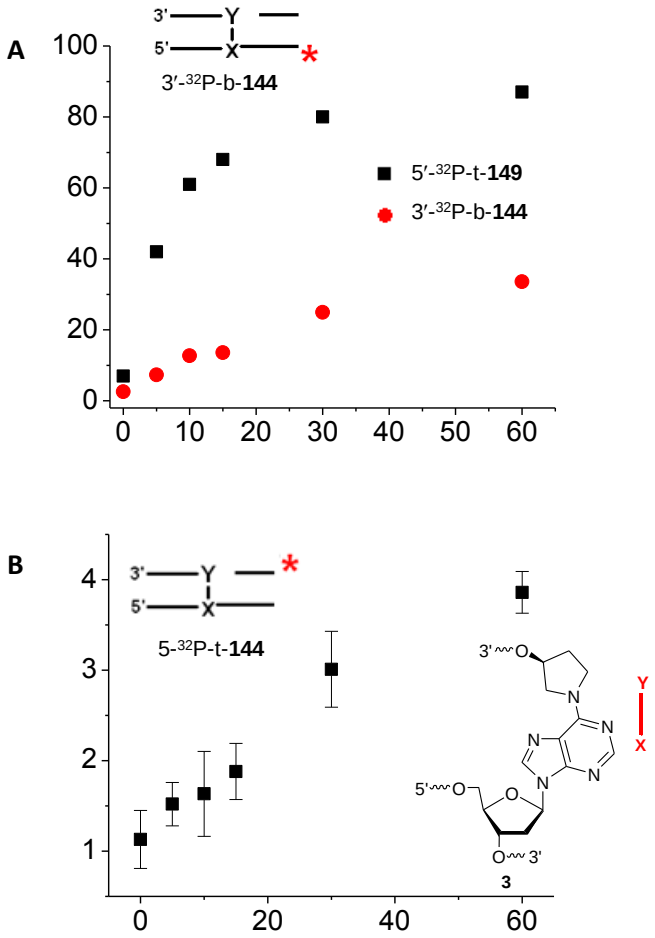


Figure 42. Time dependence of UvrABC incision of DOB ICL analogue **3** in 3′-^{32}P-b-**144** compared to that in 5′-^{32}P-t-**149** (A) and 5'-^{32}P-t-**144** (B).

After we discovered the possibility of un-cross-linked region of **144** becoming denatured under UvrABC reaction conditions, we employed oligonucleotide **146**

containing DOB ICL analogue **3** to study its incision. The incision sites of **146** resembled those of **144** and were distributed over 1 – 3 nucleotides on the 3'-side of the cross-linked A ("bottom strand", nucleotides A_{29}, G_{30} and G_{31}), albeit with a different distribution of cleavage (Figure 43 and App. Figure 37). Unlike **144**, no other incision was visible on the 5'-side of the abasic site analogue. The cleavage at dA_{29} was significantly smaller than both G_{30} and G_{31} (Figure 43 and App. Figure 37). The incision at G_{30} was the highest and slightly more than that at G_{31}. After 1 h of

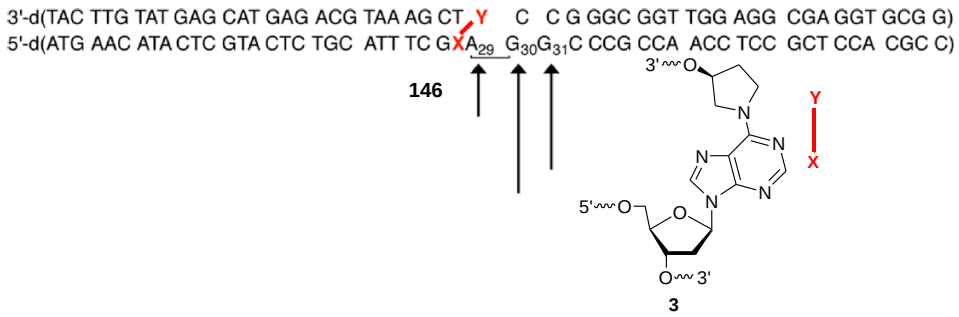


Figure 43. Histogram describing UvrABC incision of DOB ICL analogue **3** in **146**. (Data taken from 4 individual experiments in which each terminus of **146** was ^{32}P-labeled).[141]

incubation with UvrABC, the overall incision of the three nucleotides was ~16 – 22%, while 87% of the standard fluoresceinylated T (5'-^{32}P-t-**149**) was incised by that time (Figure 44). We also monitored the incubation of UvrABC with **146** when the ^{32}P-labeling was positioned at the 3'-side of the top strand (3'-^{32}P-t-**149**) or 5'-side of the bottom strand (5'-^{32}P-b-**149**). An incision between the cross-link and ^{32}P-labeling for 3'-^{32}P-t-**149** and 5'-^{32}P-b-**149** would have produced a product migrating faster than T_{81} or G_{27} from the sequencing lane, respectively (App. Figure 38 and 39). While no such product was observed for either substrate, cleavage products migrating just underneath **149** and above T_{81} (for 3'-^{32}P-t-**149**) or G_{27} (for 5'-^{32}P-b-**149**) were detected (App. Figure 38 and 39). Not only were these products consistent with incision 3' to the cross-linked A,

the yields for 5'-^{32}P-b-**146** matched those from 3'-^{32}P-b-**146** as well (App. Figure 39). In addition, treatment of 3'-^{32}P-t-**146** for 8 h with UvrABC generated more than 60% incision products (App. Figure 39). These results were consistent with the incision at A_{29} – G_{31} detected during UvrABC treatment of 3'-^{32}P-b-**149** (Figure 43 and App. Figure 38). We speculated whether the observed regioselectivity of UvrABC incision was due to the presence of the nick. During the UvrABC incision on a single strand nick, the cleavage occurs at the sixth phosphodiester bond from the strand break on the 5'-direction.[123] In contrast, the UvrABC incision on DOB ICL analogue **3** occurred on the uncleaved strand with no strand break. However, an as yet unexplored mechanism of the nick inducing the exclusive UvrABC incision on the uncleaved strand of **3** cannot be ruled out.

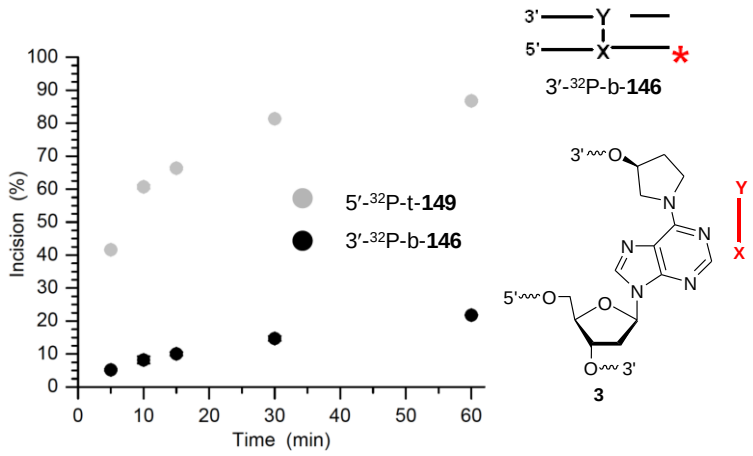


Figure 44. Time dependence of UvrABC incision of DOB ICL analogue **3** in 3′-^{32}P-b-**146** compared to that in 5′-^{32}P-t-**149**.[141]

3.2.5. Transfer of cross-linked DNA from UvrA to UvrB: preincision complex formation

After doing an initial search for damage in DNA, $UvrA_2$-DNA binds with $UvrB_2$ to perform a more thorough scan (Scheme 33).[106,110] Upon confirming the presence of a damage site, $UvrA_2$ dissociates and the $UvrB_2$-DNA complex (preincision complex) awaits the recruitment of UvrC.[110,114] Preincision complex formation is regarded as the rate-determining step of the incision.[124] Using a gel shift assay, the efficiency of $UvrA_2$-

DNA or $UvrB_2$-DNA complex formation of cross-linked oligonucleotides **143** (containing C4-AP ICL mimic **4**), **146** (containing DOB ICL mimic **3**), and **149** (containing fluoresceinylated T) was compared. In addition, the correlation between the efficiency of DNA-protein complex formation and incision yield was investigated for the three substrates. In order to investigate the $UvrA_2$-DNA complex formation by gel shift assay, aliquots from the incubation (55 °C, 20 min) of UvrA (0 – 200 nM) with DNA (0.3

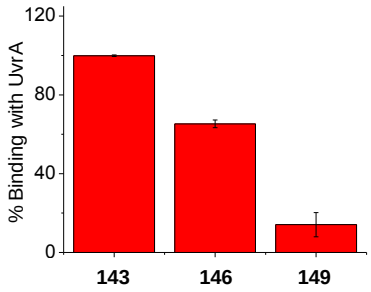


Figure 45. Comparison of binding of 0.3 nM DNA (**143**, **146** and **149**) with 80 nM DNA in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 20 min.

nM) were analyzed using 4% nondenaturing PAGE at 4 °C. The 4% nondenaturing PAGE used for the UvrAB gel shift has 80:1 acrylamide to bisacrylamide ratio to facilitate effective migration of DNA-protein complexes inside the gel.[125] Although

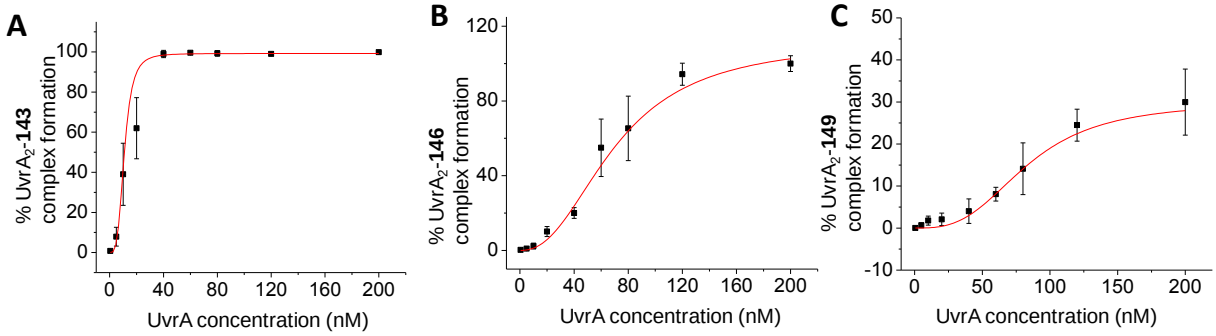


Figure 46. Binding of UvrA (0.5 – 200 nM) with 0.3 nM 3'-^{32}P-t-**143** (A), 3'-^{32}P-b-**146** (B), and 5'-^{32}P-t-**149** (C) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 20 min. Curve fitted using Hill equation for single site binding.

$UvrA_2$ complexation with 5'-^{32}P-t-**149** was weak, binding with 3'-^{32}P-t-**143**, and 3'-^{32}P-b-**146** was significantly higher (Figure 45 and App. Figure 40). Incubation with UvrA (80 nM) led to 99.8%, 65.3% and 14.1% DNA-protein complex formation with 3'-^{32}P-t-**143**, 3'-^{32}P-b-**146**, and 5'-^{32}P-t-**149** respectively (Figure 46 and Table 1). The dissociation constant or K_d is defined as the protein concentration at which half of the DNA is in the form of DNA-protein complex. The K_d for $UvrA_2$ binding with 3'-^{32}P-t-**143**, 3'-^{32}P-b-**146**, and 5'-^{32}P-t-**149** was calculated by fitting the $UvrA_2$-DNA-vs UvrA binding isotherm with the Hill equation[142,143] for single site binding (Figure 46). The K_d for **143** and **146** matched the previously reported numbers of UvrA binding of single stranded lesions, while the K_d for 5'-^{32}P-t-**149** is slightly higher.[112,113] The weak binding noticed for $UvrA_2$-DNA complexes is consistent with the catalytic nature of UvrA.

Table 1. Binding of damaged DNA with UvrA (considering dimer formation)

Substrate	Binding (%) with 80 nM UvrA	Dissociation Constant (K_d, nM)
3'-^{32}P-t-**143**	99.87 ± 0.38	5.45 ± 1.10
3'-^{32}P-t-**146**	65.30 ± 1.98	22.13 ± 5.02
5'-^{32}P-t-**149**	14.11 ± 6.14	32.98 ± 10.31

Next, we employed gel shift assays to probe the rate determining step of UvrABC incision, formation of the $UvrB_2$-DNA preincision complex following transfer from $UvrA_2$.[124] We anticipated that the observed incision yields could be interpreted based on the efficiency of preincision complex formation for 3'-^{32}P-t-**143**, 3'-^{32}P-b-**146**, and 5'-^{32}P-t-**149**. Aliquots from incubation of DNA (0.3 nM), UvrB (500 nM), and UvrA (1, 2, 4, 8 and 10 nM) were analyzed by gel shift to determine the catalytic UvrA concentrations required for effective transfer to UvrB (App. Figure 41 – 43). The high UvrB concentration was to drive the equilibrium towards $UvrA_2$ dissociation and $UvrB_2$-DNA

preincision complex formation. The transfer of 3'-^{32}P-t-**143**, 3'-^{32}P-b-**146**, and 5'-^{32}P-t-**149** from $UvrA_2$ to $UvrB_2$ displayed a trend which was completely

Table 2. Hand-off of damaged DNA from UvrA (specified concentration) to UvrB (500 nM)

UvrA concentration (nM)	Cross-linked substrate 3'-^{32}P-t-**143** containing C4-AP ICL analogue **4** binding (%)	Cross-linked substrate 3'-^{32}P-t-**146** containing DOB ICL analogue **3** binding (%)	5'-^{32}P-t-**149** containing fluoresceinylated T binding (%)
1	4.62 ± 1.27	5.52 ± 0.05	31.56 ± 3.69
2	5.67 ± 2.08	6.03 ± 0.78	60.03 ± 4.42
4	11.89 ± 2.09	11.42 ± 0.28	74.46 ± 1.80
8	18.94 ± 2.67	21.29 ± 1.88	87.33 ± 0.23
10	23.89 ± 3.08	24.22 ± 2.09	86.37 ± 2.07

opposite to that of $UvrA_2$ binding. Strong binding with UvrB (> 60%) was detected for 5'-^{32}P-t-**149** for UvrA concentrations as small as 2 nM, whereas modest binding (> 10%) was possible for **143** and **146** only when the UvrA was present in 4 and 8 nM (Figure 47, App. Figure 41 – 43, and Table 2). As a result, incubation of **143**, **146** and **149** (0.3 nM) with UvrA (8 nM) and UvrB (500 nM) led to 18.9%, 21.3% and 87.3% preincision complex formation (Figure 47 and Table 2). The greater efficiency of preincision complex formation for **149** was fully consistent with its higher incision yield. However, UvrB binding with **143** and **146** with similar efficiency did not correlate with the relatively greater incision observed for **146** (Figure 47 and Table 2). We hypothesize this could be due to significantly high UvrB concentration (500 nM) used for the transfer assays, that might have diluted the selectivity between **143** and **146**. Independent incubation of mitomycin C ICL (0.6 – 2 nM) with UvrA alone (15 nM) or with UvrA and UvrB (15 nM each) led to almost quantitative protein-DNA complex formation.[20] Mitomycin C ICL formed a near-quantitative $UvrB_2$-DNA complex (preincision

complex) formation. Despite this, the reported gels indicted that the incision yield for mitomycin C was less than 50% (no numbers were provided in the paper).[20] Although a

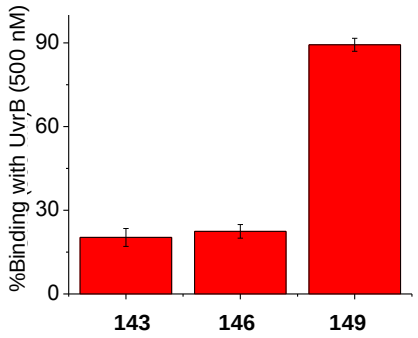
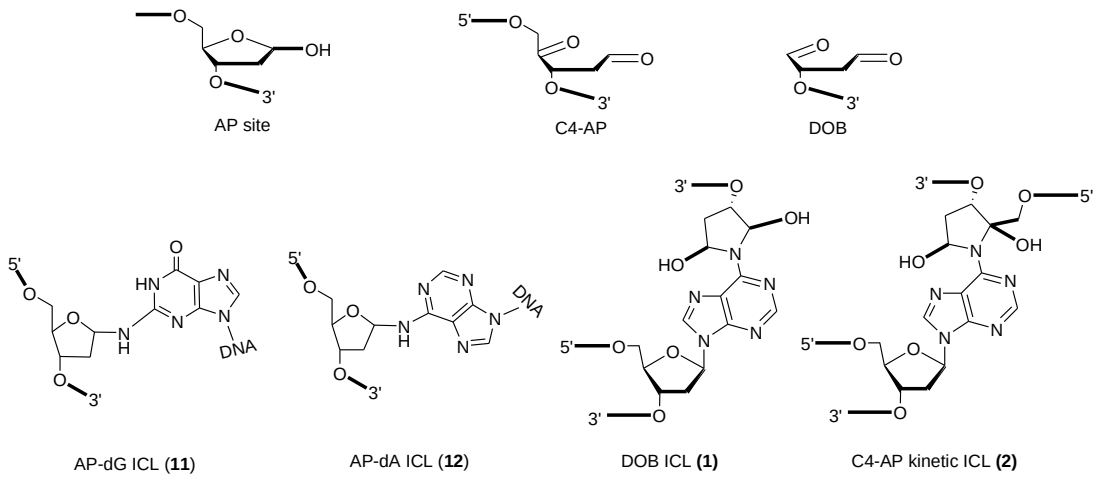


Figure 47. Binding of 3'-^{32}P-t-**143**, **3'**-^{32}P-t-**146**, and 5'-^{32}P-t-**149** (concentration 0.3 nM) with 500 nM UvrB following transfer from UvrA (8 nM) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 20 min.

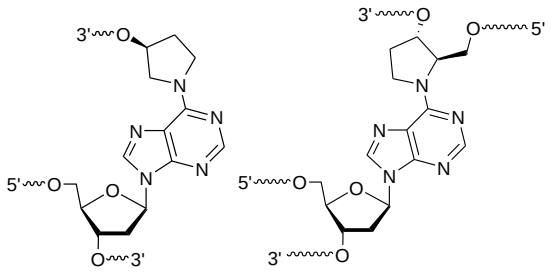
dissociation constant was not computed for the strong UvrA binding of mitomycin C ICL, it resembled the stronger cross-link affinity seen for **143** and **146**.[20] The stronger affinity of UvrA towards cross-linked substrates is ambiguous from standpoint of the current understanding of UvrABC incision model and needs further research.

3.2.6. Summary



Agents causing interstrand cross-links are widely considered to have the highest genotoxicity.[103] ICLs are capable of blocking essential cellular processes like

transcription[11] and replication[10] by preventing strand separation. Molecules that produce cross-links are of the utmost interest in terms of developing therapeutic agents, as well as for recognizing potential ICL producing conditions. Recently, even simple abasic sites (AP, C4-AP, and DOB) have been implicated in cross-link formation.[6,7,9,31] While the cross-links produced from C4-AP and DOB (**1** and **2**) have short life spans, they are still capable of disrupting cellular processes.



In this study, we treated stabilized analogues of DOB and C4-AP ICLs (**3** and **4**) with UvrABC, the bacterial nucleotide excision repair proteins, to gain insight into their repair. The cross-link analogues are synthesized by solid phase synthesis and enzymatic ligation. Treatment of the DOB ICL analogue (**3**) with UvrABC exclusively produced double strand breaks (DSBs), which is the first example of an UvrABC repair gone completely wrong. The yield of the DSB is as high as 60% in 8 h. Treatment of C4-AP ICL analogue **4** with UvrABC revealed that it is a poor substrate. Since ICL removal is necessary to resume blocked activities like replication and transcription in a cell, persistence of an ICL against repair proteins can have a very deleterious effect on cellular processes. In the gel shift assay binding of cross-links with UvrA and UvrB, two damage recognition proteins, was investigated. We experienced a slightly stronger affinity of UvrA towards the cross-links **3** and **4** that a standard substrate (**149**) containing

fluoresceinylated T. The poor efficiency of preincision complex formation explained the low incision yield observed for **3** and **4**. Consequently, this study based on stabilized analogues of **1** and **2** addressed the biochemical effects of their formation.

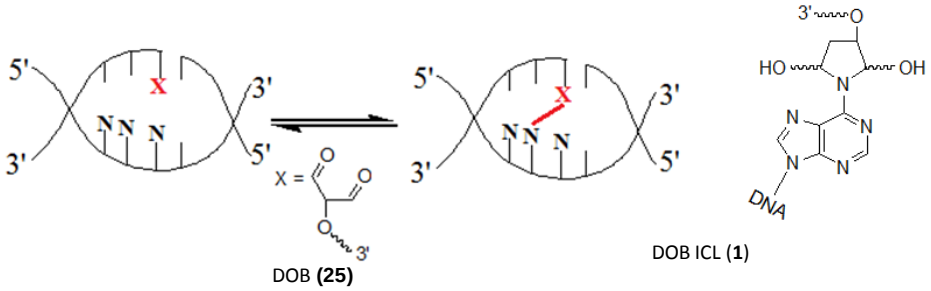
3.3. Structural probing of ICL

3.3.1. Structural probing of DOB ICL mimics

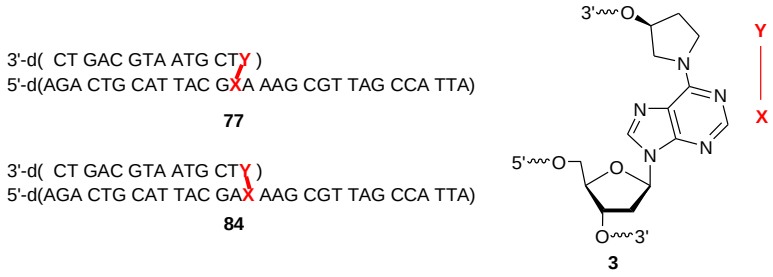
3.3.1.1. Chemical probing of cross-linked DNA containing DOB ICL analogue 3

3.3.1.1.1. Chemical probing reactions and substrate design

Scheme 11. ICL generation by DOB lesion



During the formation of DOB interstrand cross-link (**1**), the electrophilic lesion (**25**) selectively formed an adduct with N6 of an dA opposite the 3'-adjacent nucleotide (Scheme 11).[9] Compared to 21 – 38% yield for ICL formation with dA opposite 3'-adjacent nucleotide, only 1 – 3.5% adduct formation was observed with the opposing dA.[9] The N6 amine of the dA opposite the 3'-adjacent nucleotide was less than 0.4 Å



closer to the C1' aldehyde of DOB than that of an opposing A. The site selectivity thus cannot be justified based on the proximity alone. We considered whether the instability of the transition state during reversible formation of **1** with an opposing A could be another factor. We hypothesized that the instability originates from structural distortion associated with forming **1** with the opposing A. To test the hypothesis, two cross-link substrates containing DOB ICL analogue **3** were constructed. Substrate **77**, consisting of a cross-link with A opposite the 3'-adjacent nucleotide mimicked the "observed" ICL formation. On the other hand, in **84** cross-

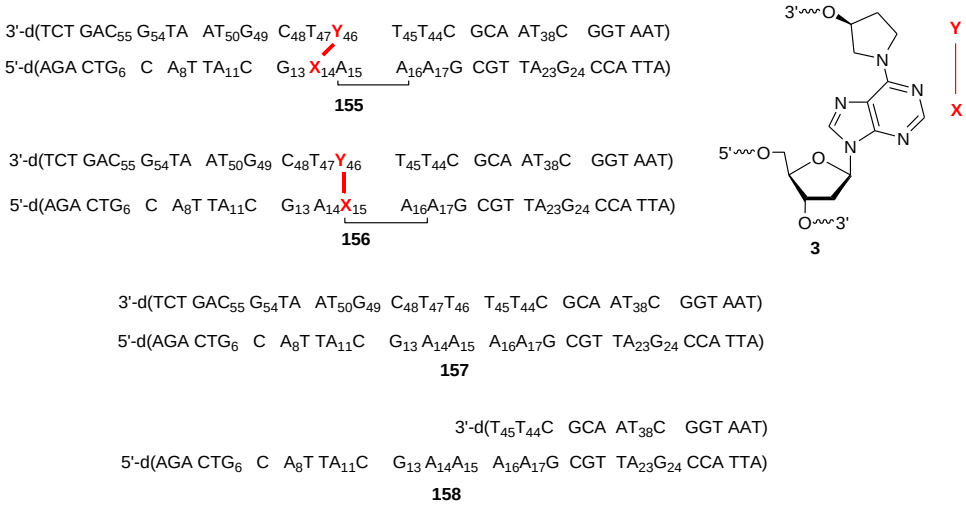


Figure 48. Oligonucleotides containing **3** and standard substrates used for chemical agent based structural probing.

linking occurs with an opposing A, and mimics the "unobserved" (less observed) ICL. The region around the cross-linked A was kept A rich to mimic the flanking sequences of the native DOB ICL (**1**).[9] The strands containing oligonucleotides **77** and **84** were converted to **155** and **156**, containing fully complementary strands (Figure 48) and were probed using chemical agents. See Experimentals for details regarding the construction of **155** and **156**. Nucleotide specific chemical agents used for structural probing are based on the principles of Maxam-Gilbert sequencing experiments.[144] The reagents used in this

study included hydroxyl radical cleaving the sugar backbone, diethyl pyrocarbonate (DEPC, final concentration 1.04 M) for A, dimethyl sulfate (DMS, final concentration 25 mM) for G, hydroxylamine (final concentration 0.8 M) for C, and $KMnO_4$ (final concentration 0.3 mM) for T. These reagents preferentially react with solvent exposed distorted nucleotides in a strand to create alkali-labile sites (except for hydroxyl radical cleavage, which cleaves the backbone without any alkali treatment). As a result, nucleotides in bulges, overhangs, loops and distorted regions are more vulnerable to these reagents than regular Watson-Crick base pairs. The base-labile sites are then cleaved upon piperidine treatment. If the treated strand is ^{32}P-labeled at the terminus, the cleavage sites can be visualized on denaturing PAGE, pinpointing distortion in a DNA sequence. Hydroxyl radical cleavage mediated by Fe(II) based Fenton reaction is widely used to quantitatively predict DNA backbone solvent accessibility.[145,146] $KMnO_4$ and hydroxylamine are usually employed to detect point mutations in DNA.[147,148] DEPC and DMS chemical probing have been applied to detect bulged nucleotides present in loops,[149] G-quadruplexes[150] as well as in solvent exposed surfaces formed as a result of strand invasion.[151]

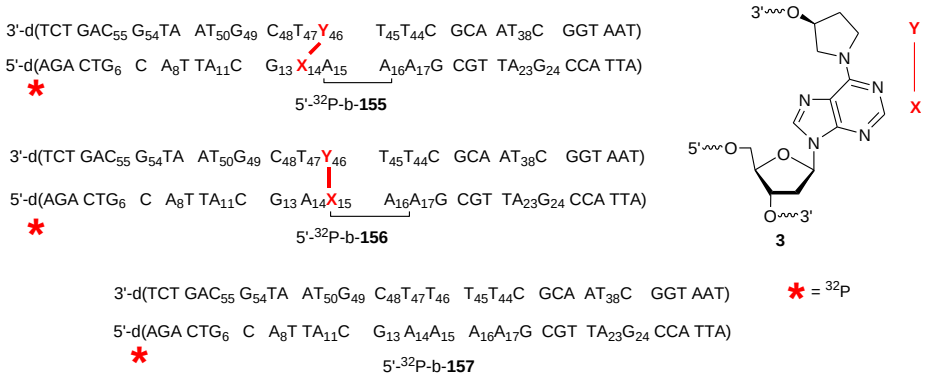
Therefore, when a terminus in **155** or **156** is ^{32}P-labeled, treated with chemical agents followed by piperidine treatment, sites of distortion between the label and cross-link are revealed as strand breaks. Each terminus of **155** and **156** was thus independently labeled and treated with chemical agents to detect solvent exposed nucleotides between the respective terminus and cross-link. Following ^{32}P-labeling of **155** and **156**, they were purified by denaturing PAGE to ensure greater purity. The concentration of **155** and **156** during exposure to chemical probing reagents ranged between 30 – 60 nM. We pursued

single hit kinetics (more than 80% of the DNA would be unaffected by the reaction) to ensure the DNA molecule reacted with the chemical agent only once on average.[152,153] The single hit kinetics conditions was expected to maximize the sensitivity of solvent exposed sites towards chemical agents. From earlier reports on cross-link structures[92] we speculated that increased solvent exposure caused by distortion would be limited to within three or four nucleotides from the cross-link site in each strand. Under this pretext, the ratio of cleavage band volume of a nucleotide immediately adjacent to the ICL to that of a distant nucleotide (preferably the same one and at least five nucleotides away) will serve as a measure of distortion. This ratio will act as an internal normalization to separate distortion based reactivity ("hyperreactivity") from any variation in exposure to the chemical reagent. We define this ratio as "normalized cleavage" [NC = $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$], which can detect genuine solvent exposed nucleotides or "hyperreactive sites". However, solvent exposed nucleotides can arise due to local sequence effects as well.[154,155] To avoid misinterpreting any sequence derived "hyperreactivity" as a cross-link influenced one, the NC value of a nucleotide was compared with a comparable one in the duplex (**157**) without a cross-link (Figure 48). Oligonucleotide **158** (Figure 48) was used as an additional control for investigating presence of distortions in nucleotides 5' to T_{45} in **155** and **156**. The two step selection process, cleavage normalization (division by the cleavage band volume of an undistorted nucleotide) and screening (exclusion of sequence based hyperreactivity using **157** and **158**) would detect the ICL induced hyperreactive sites in **155** and **156**. Each experiment was replicated at least six times and NC values were subjected to Student's t test to verify their statistical significance (App. Table 2 – 5). Comparing the hyperreactivity of

nucleotides in **155** and **156** would thus help detect their structural distortion, and perhaps provide a rationalization for the observed selective formation of the former.

3.3.1.1.2. Investigation of hyperreactive sites at 5'-side of the bottom strand (cross-linked A containing strand)

In the region 5' to the cross-linked A (X_{14} in **155** and X_{15} in unobserved ICL **156**), the reactivity of the nucleotides A_{11} to A_{14} were investigated using the normalized cleavage or NC value [$\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}$]. The cross-linked A X_{14} (observed ICL 5'-^{32}P-b-**155**) and X_{15} (unobserved ICL 5'-^{32}P-b-**156)** were detected during the A-G sequencing (formic acid, room temperature) of the respective substrates



(Lanes 12 and 18, App. Figure 44A). During the treatment of 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156** and 5'-^{32}P-b-**157** [the strand containing the abasic site analogue (T_{46} for **157**) was referred to "top" strand or "t" in each substrate, while the cross-linked A (A_{14} and A_{15} for **157**) containing strand was named as "bottom" or "b"] with hydroxyl radical, A_8 (five nucleotides away from cross-linked A) was used as the reference nucleotide. The histograms for the three substrates obtained from denaturing PAGE analysis did not reveal significant hyperreactivity for cross-linked X_{14} (for observed 5'-^{32}P-b-**155)** or X_{15} (for unobserved 5'-^{32}P-b-**156**) compared to the respective nucleotides in 5'-^{32}P-b-**157**

(App. Figure 43). However, comparison of the normalized cleavage ("NC" values) indicated that the reactivity at cross-linked position (X_{14}) was higher in 5'-^{32}P-b-**155** (observed ICL, NC = 0.94 ± 0.07) compared to that (X_{15}) in 5'-^{32}P-b-**156** (unobserved ICL, NC = 0.75 ± 0.04) (Figure 49A and Table 3).

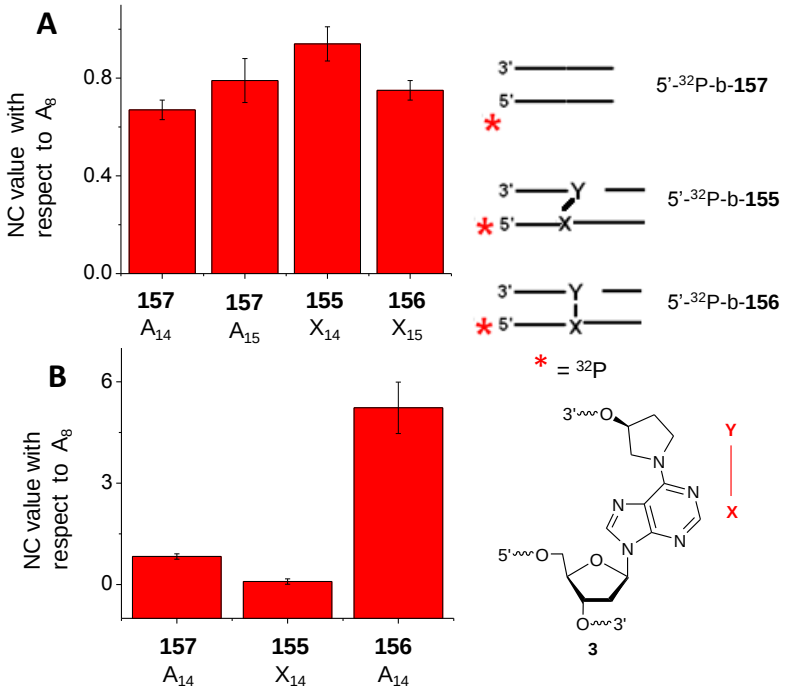


Figure 49. Comparison of NC value of nucleotides (as a function of distance from ICL) in 5'-^{32}P-b-**155** (observed ICL), 5'-^{32}P-b-**156** (unobserved ICL), and 5'-^{32}P-b-**157** (control duplex) for hydroxyl radical cleavage and DEPC reaction. (A) NC value of hydroxyl radical reaction of cross-linked A from 5'-^{32}P-b-**155**, and 5'-^{32}P-b-**156** and A_{14}, A_{15} from 5'-^{32}P-b-**157** with respect to A_8. (B) NC value of DEPC reaction of A_{14} from 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156** and 5'-^{32}P-b-**157** with respect to A_8.

The diethylpyrocarbonate (DEPC) reaction on 5'-^{32}P-b-**155** (observed ICL), 5'-^{32}P-b-**156** (unobserved ICL) and 5'-^{32}P-b-**157** (control duplex) did not react at the cross-linked As (X_{14} and X_{15} in 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156**, respectively). The NC values $[\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$, reference nucleotide A_8] indicated that A_{14} in 5'-^{32}P-b-**156** (unobserved ICL, NC = 5.22 ± 0.77) was higher than the respective nucleotide in 5'-^{32}P-b-**157** (control duplex, NC = 0.82 ± 0.07) (App Figure 45, Figure

Table 3. Chemical probing experiments on DOB ICL analogue 3 on the 5'-side of A containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
A_{14} of **157** A_{15} of **157**	0.67 ± 0.04 0.79 ± 0.09	A_8 (hydroxyl radical cleavage)
Cross-linked position of **155** (X_{14})	0.94 ± 0.07	
Cross-linked position of **156** (X_{15})	0.75 ± 0.04	
A_{14} of **157**	0.82 ± 0.07	A_8 (DEPC)
X_{14} of **155**	0.09 ± 0.08	
A_{14} of **156**	5.22 ± 0.77	
G_{13} of **157**	0.81 ± 0.08	G_6 (DMS)
G_{13} of **155**	1.25 ± 0.07	
G_{13} of **156**	1.41 ± 0.07	

49B and Table 3). This indicated loss of H-bonding at A_{14} in 5'-^{32}P-b-**156** with the opposing T_{47} caused by ICL induced distortion. The cleavage intensity of A_{14} was greater than that of A_8 in unobserved ICL 5'-^{32}P-b-**156** (App. Figure 45D). In contrast, the cleavage intensities of the same nucleotides in 5'-^{32}P-b-**157** were similar (App. Figure 45B). This indicates how reactivity (therefore distortion) decreases as one moves further from the cross-link.

The result of dimethyl sulfate (DMS) treatment of 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156** and 5'-^{32}P-b-**157** was more informative than the DEPC reaction. The reactivity of G_{13} (one nucleotide away from cross-link) was greater than G_6 (7 nucleotides away from cross-link) in 5'-^{32}P-b-**155** (observed ICL) and 5'-^{32}P-b-**156** (unobserved ICL) but not in 5'-^{32}P-b-**156** (control duplex) (App. Figure 46). This reaffirmed the presence of distortion in nucleotides around the DOB ICL. The NC value

$[\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}$, reference nucleotide G_6] for G_{13} was the highest for unobserved ICL (5'-^{32}P-b-**156**, NC = 1.41 ± 0.07) among the three substrates (Figure 50A and Table 3).

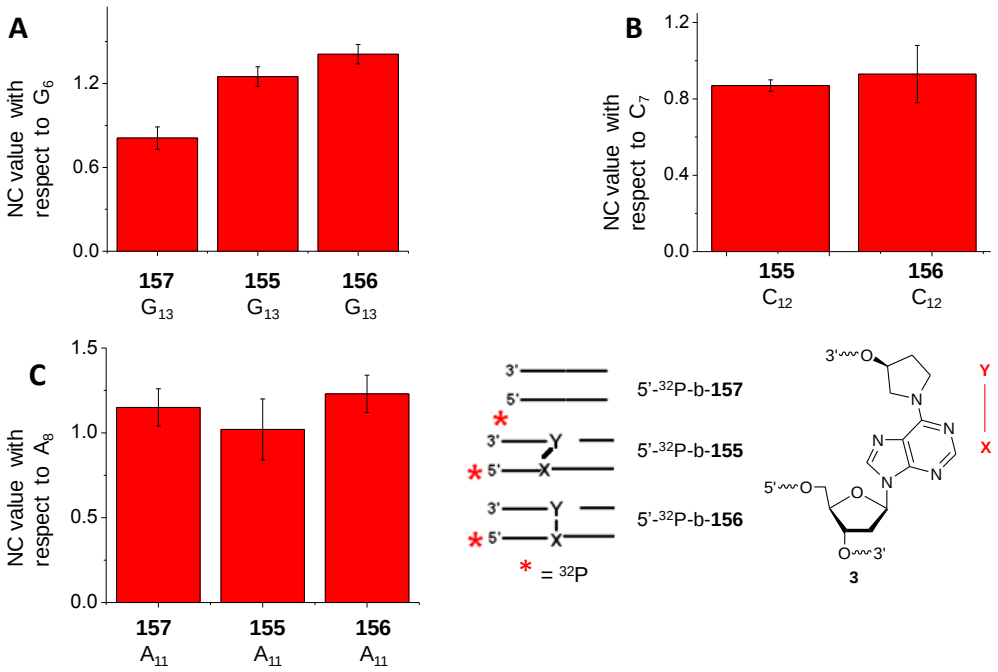
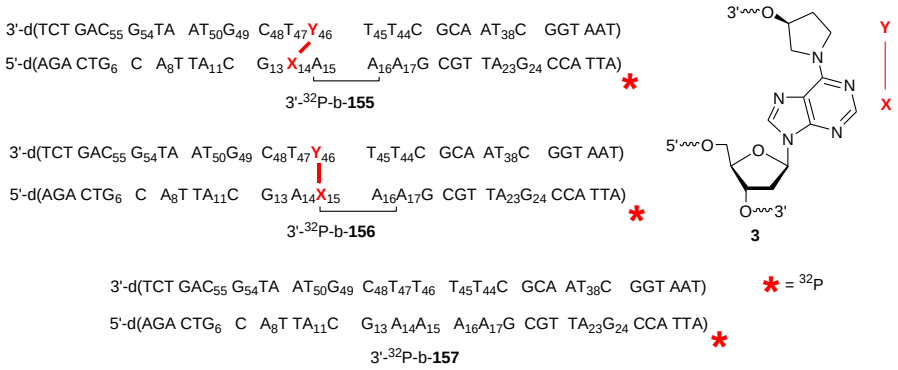


Figure 50. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 5'-^{32}P-b-**155** (observed ICL), 5'-^{32}P-b-**156** (unobserved ICL), and 5'-^{32}P-b-**157** (control duplex) for DMS, hydroxylamine and DEPC reactions. (A) NC value of DMS reaction of G_{13} from 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156**, and 5'-^{32}P-b-**157** with respect to G_6. (B) NC value for hydroxylamine reaction of C_{12} from 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156** and 5'-^{32}P-b-**157** with respect C_7. (C) NC value for DEPC reaction of A_{11} from 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156** and 5'-^{32}P-b-**157** with respect to A_8.

We probed C_{12} for distortion using reaction with hydroxylamine. However, as shown here for C_{12} and other C nucleotides around the ICL, the hydroxylamine reaction used to detect distortions at C nucleotide was less effective than other chemical probing agents. Since it did not demonstrate a high degree of selectivity between cross-link adjacent C_{12} (2 nucleotide away from ICL) and reference nucleotide C_7 (6 nucleotide away from ICL), it did not reproducibly indicate hyperreactivity at C_{12} (App. Figure 47). Although the NC for C_{12} in unobserved ICL 5'-^{32}P-b-**156** (NC = 0.93 ± 0.15) had a higher

value than in the observed ICL 5'-^{32}P-b-**155** (NC = 0.93 ± 0.15) (indicating greater distortion), the lack of reproducibility was associated with a high standard deviation (Figure 50B and App. Table 1). Finally, NC values for the next nucleotide A_{11} (3 nucleotides away from ICL) with reference to A_8 were calculated based on the same DEPC reaction employed to determine the hyperreactivity at A_{14} (App. Figure 45). Although the NC was higher for 5'-^{32}P-b-**156** (unobserved ICL), the difference between 5'-^{32}P-b-**155** (observed ICL) and 5'-^{32}P-b-**156** for A_{11} were statistically insignificant (Figure 50A, and App. Table 1). Interpreting this as an indication of dissipating distortion, reactivity of any other nucleotides in this region was not pursued any further.

3.3.1.1.3. Investigation of hyperreactive sites on the 3'-side of the bottom strand (cross-linked A containing strand)



In the region 3' to the cross-linked A (X_{14} in observed ICL **155** and X_{15} in unobserved ICL **156**), the reactivity of the nucleotides A_{15} to A_{17} were investigated using their normalized cleavage or NC value $[\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}]$. The general trend of greater hyperreactivity in 3'-^{32}P-b-**156** (unobserved ICL) compared to 3'-^{32}P-b-**155** (observed ICL) matched that seen in 5'-^{32}P-b-**156** and 3'-^{32}P-b-**155**. Since the nucleotides on the 3'-side of the cross-linked A containing strand proximal to the modification primarily consisted of As, 3'-^{32}P-b-**155**, 3'-^{32}P-b-**156**, and 3'-^{32}P-b-**157** were

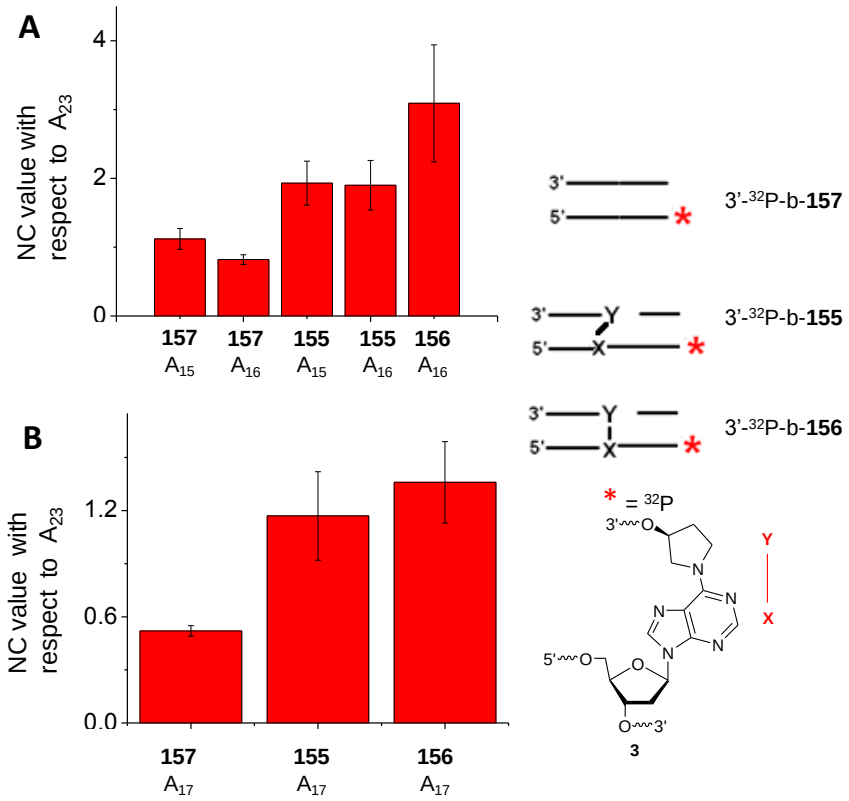


Figure 51. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 3'-^{32}P-b-**155** (observed ICL), 3'-^{32}P-b-**156** (unobserved ICL), and 3'-^{32}P-b-**157** (control duplex) for DEPC reaction. (A) NC value of DEPC reaction of nucleotides from 3'-^{32}P-b-**155** (A_{15} and A_{16}), 3'-^{32}P-b-**156** (A_{16}), and 3'-^{32}P-b-**157** (A_{15} and A_{16}) with respect to A_{23}. (B) NC value of DEPC reaction of A_{17} from 3'-^{32}P-b-**155** , 3'-^{32}P-b-**156**, and 3'-^{32}P-b-**157** with respect to A_{23}.

treated with DEPC. Although the cross-linked A (X_{14} for 3'-^{32}P-b-**155** and X_{15} for 3'-^{32}P-b-**156**) again did not react with DEPC, a number of 3'-adjacent As were hyperreactive (App. Figure 48). First, we compared the reactivity of cross-link adjacent positions, A_{15} in observed ICL (3'-^{32}P-b-**155**) and A_{16} in unobserved ICL (3'-^{32}P-b-**156**). The NC value of A_{16} (reference nucleotide A_{23}) in unobserved ICL (3'-^{32}P-b-**155**, NC = 3.09 ± 0.85) was greater than A_{15} in observed ICL (3'-^{32}P-b-**155**, NC = 1.93 ± 0.32) (Table 4 and Figure 51A), indicating greater distortion. This observation was intriguing since A_{16} in 3'-^{32}P-b-**156** was opposite T_{45} (capable of H-bnding) while A_{15} in 3'-^{32}P-b-**155** was opposite an abasic site (completely devoid of H-bonding). In addition, the NC value [NC =

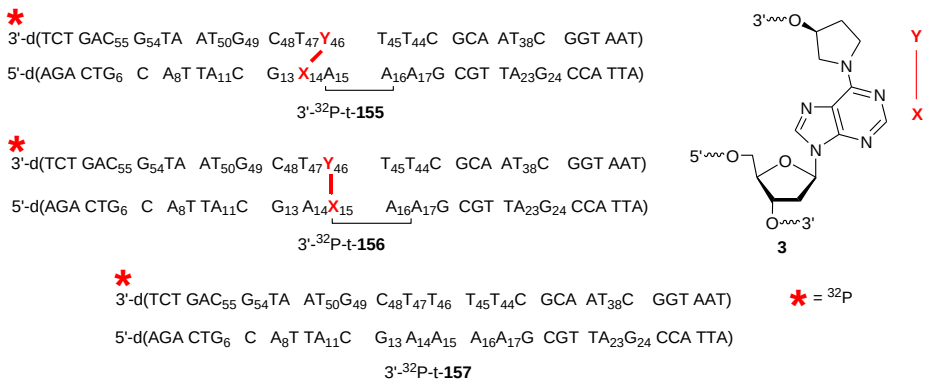
$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$, reference nucleotide A_{23}] of A_{16} (one nucleotide away from ICL) in 3'-^{32}P-b-**156** (unobserved ICL, NC = 3.09 ± 0.85) was

Table 4. Chemical probing experiments on DOB ICL analogue 3 on the 3'-side of A containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
A_{15} of **157** A_{16} of **157**	1.12±0.15 0.82 ± 0.07	A_{23} (DEPC)
Position adjacent to cross-link in **155** (A_{15})	1.93 ± 0.32	
Position adjacent to cross-link in **156** (A_{16})	3.09 ± 0.85	
A_{16} of **155**	1.90 ± 0.36	A_{23} (DEPC)
A_{16} of **156**	3.09 ± 0.85	
A_{17} of **157**	0.52±0.03	A_{23} (DEPC)
A_{17} of **155**	1.17 ± 0.25	
A_{17} of **156**	1.36 ± 0.23	

higher than (A_{16}) that in 3'-^{32}P-b-**155** (observed ICL, NC = 1.90 ± 0.36) (Figure 51A and Table 4), signifying greater distortion for 3'-^{32}P-b-**156**. The A_{17} (two nucleotides away from ICL) in 3'-^{32}P-b-**156** has a higher NC value (reference nucleotide A_{23}) than in 3'-^{32}P-b-**155** and in 3'-^{32}P-b-**157**, but it is within statistical error (Figure 51B and Table 4), indicating gradually diminished influence of distortion. Nucleotides 3'- to A_{17} were not probed. The general trend of hyperreactivity in 3'-^{32}P-b-**156** matched that in 5'-^{32}P-b-**156**, indicating greater distortion for the "unobserved" in which DOB is cross-linked with the opposing base.

3.3.1.1.4. Investigation of hyperreactive sites on the 3'-side of the top strand (abasic site analogue containing strand)



In the region 3' to the cross-linked abasic site (Y_{46} in **155** and **156**), the reactivity of nucleotides Y_{46} to T_{50} were investigated using their normalized cleavage or NC value $[\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}]$. Hydroxyl radical cleavage on the 3'-side of

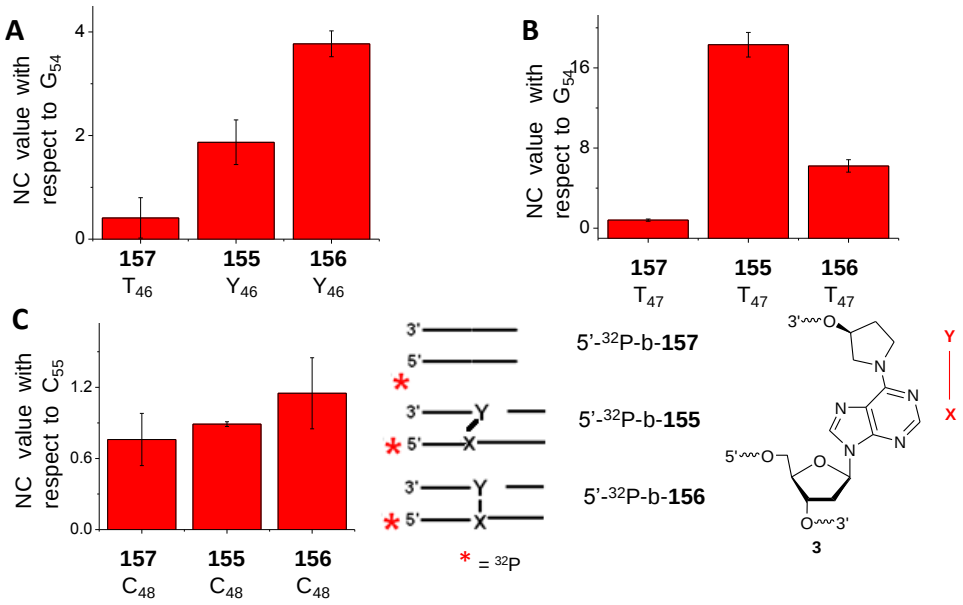


Figure 52. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 3'-^{32}P-t-**155** (observed ICL), 3'-^{32}P-t-**156** (unobserved ICL), and 3'-^{32}P-t-**157** (control duplex) for hydroxyl radical cleavage, $KMnO_4$ and hydroxylamine reaction. (A) NC value for hydroxyl radical reaction of T_{46} (Y_{46} for cross-links) from 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156** and 3'-^{32}P-t-**157** with reference to G_{54}. (B) NC value for $KMnO_4$ reaction of T_{47} from 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, 3'-^{32}P-t-**157** with respect G_{54}. (C) NC value for hydroxylamine reaction of C_{48} from 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, 3'-^{32}P-t-**157** with respect C_{55}.

the cross-linked abasic site ("top" strand) indicated that the reactivity at cross-linked position Y_{46} in 3'-^{32}P-t-**155** and 3'-^{32}P-t-**156** is significantly higher than the corresponding un-cross-linked T_{45} in 3'-^{32}P-t-**157** (Figure 52A). In fact, the high reactivity of the abasic

Table 5. Chemical probing experiments on DOB ICL analogue 3 on the 3'-side of abasic side analog containing strand ("top" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
T_{46} of **157**	0.41 ± 0.39	G_{54} (hydroxyl radical cleavage)
Cross-linked position of **155** (Y_{46})	1.87 ± 0.43	
Cross-linked position of **156** (Y_{46})	3.77 ± 0.25	
T_{47} of **157**	0.81 ± 0.11	G_{54} ($KMnO_4$)
T_{47} of **155**	18.31 ± 1.23	
T_{47} of **156**	6.22 ± 0.62	
G_{49} of **157**	1.56 ± 0.02	G_{54} (DMS)
G_{49} of **155**	1.45 ± 0.22	
G_{49} of **156**	1.90 ± 0.04	
T_{50} of **157**	1.84 ± 0.12	G_{54} ($KMnO_4$)
T_{50} of **155**	1.48 ± 0.08	
T_{50} of **156**	2.14 ± 0.27	

site analogue in 3'-^{32}P-t-**156** produced a prominent band in the gel (App. Figure 49). The high reactivity of the pyrrolidine ring could be from greater solvent exposure itself or inherently higher reactivity of the pyrrolidine ring. The NC value (3.77 ± 0.25, reference nucleotide G_{54}) for the cross-linked position in the top strand (Y_{45}) of 3'-^{32}P-t-**156** (unobserved ICL) was two times greater than that for 3'-^{32}P-t-**155** (observed ICL, NC = 1.87 ± 0.43) (Figure 52A and Table 5).

$KMnO_4$ reaction for the same substrates however showed an opposite trend. For duplex 3'-^{32}P-t-**157**, almost no reaction was observed, indicating low reactivity of the based paired Ts towards $KMnO_4$ treatment (App. Figure 50A). The reactivity of T_{47} 3'-

adjacent to the cross-link in the top strand was higher for 3'-^{32}P-t-**155** than that in 3'-^{32}P-t-**156** (App. Figure 50A). The NC value [$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$, reference nucleotide G_{54}] for T_{47} in 3'-^{32}P-t-**155** (observed ICL, NC = 18.31 ± 1.23)

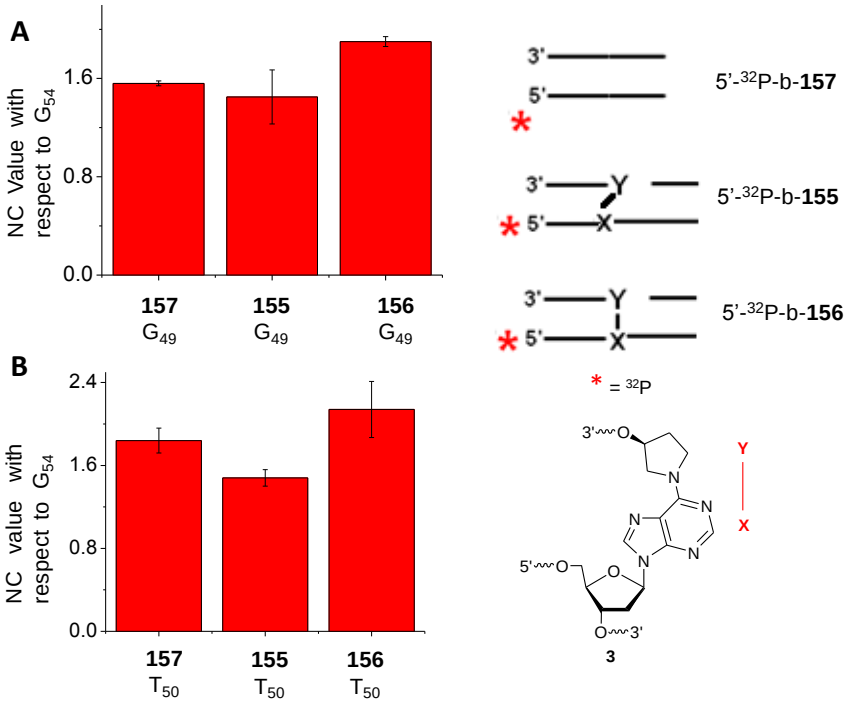


Figure 53. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 3'-^{32}P-t-**155** (observed ICL), 3'-^{32}P-t-**156** (unobserved ICL), and 3'-^{32}P-t-**157** (control duplex) for DMS and $KMnO_4$ reaction. (A) NC value for DMS reaction of G_{49} from 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, and 3'-^{32}P-t-**157** with reference to G_{54}. (B) NC value for $KMnO_4$ reaction of T_{50} from 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156** and 5'-^{32}P-b-**157**.with respect to G_{54}.

was almost three times that of 3'-^{32}P-t-**156** (unobserved ICL, NC = 6.22 ± 0.62) (Figure 52B and Table 5). The enigmatic hyperreactivity of T_{47} in the observed ICL (3'-^{32}P-t-**155**) was attributed to the presence of the opposing cross-linked A (X_{14}) and resulting disruption of duplex structure. Interestingly, T_{47} in the top strand of 3'-^{32}P-t-**156** showed significant reactivity, despite being base-paired with A_{14} (App. Figure 50A and D). This effect was consistent with the hyperreactivity detected for A_{14} in 5'-^{32}P-b-**156** and was interpreted to be a consequence of ICL induced distortion. An intriguing observation is

the presence of a strong band right above that of T_{45} for cross-linked substrates 3'-^{32}P-t-**155** and 3'-^{32}P-t-**156** (App. Figure 50A). It co-migrated with the cross-linked position Y_{45}, an abasic site. Since no reactivity is expected of an abasic site towards $KMnO_4$, its origin remains ambiguous. The reactivity of the next nucleotide of 3'-^{32}P-t-**155** and 3'-^{32}P-t-**156** (C_{48}) was more ambiguous.

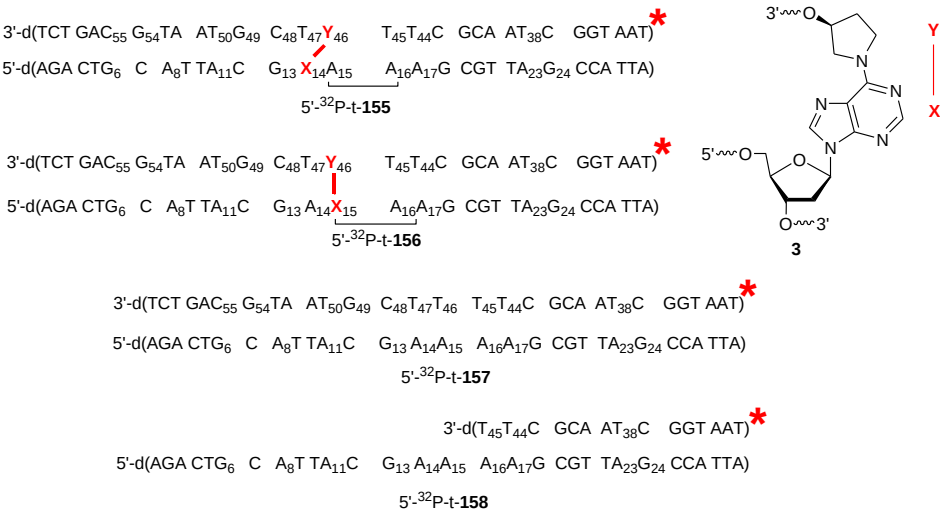
The peak volume in the histograms revealed that the reactivity of the C_{48} (two nucleotides away from ICL) was generally higher than C_{55} (nine nucleotides away from ICL) in the cross-linked substrates 3'-^{32}P-t-**155** and 3'-^{32}P-t-**156**, but very similar for duplex 5'-^{32}P-t-**157**, indicating presence of distortion in the ICL substrates (App. Figure 51). While the average NC value of C_{48} was higher for 3'-^{32}P-t-**156** (unobserved ICL, NC = 1.15 ± 0.30) than 3'-^{32}P-t-**155** (observed ICL, NC = 0.89 ± 0.02), the standard deviation associated with the former was high, making a definite conclusion impossible (Figure 52C and App. Table 3). In fact, the higher uncertainties associated with the hydroxylamine reaction was a recurring observation in the course of the chemical probing reactions.

In contrast, the DMS reaction proved to be more reliable and reproducible in terms of detecting hyperreactivity. Treatment of duplex 5'-^{32}P-t-**157** with DMS yielded higher cleavage at G_{49} compared to G_{54} (eight nucleotides away from ICL), revealing an inherently higher solvent exposure present in the former (App. Figure 52A and B). G_{49} in cross-linked substrates 5'-^{32}P-t-**155** and 5'-^{32}P-t-**156** was even more reactive than its peer in the control duplex (5'-^{32}P-t-**157**) (App. Figure 52A, C and D). The NC value of G_{49} [$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$, reference nucleotide G_{54}] in unobserved

ICL (5'-^{32}P-t-**156**, NC = 1.90 ± 0.04) was higher than that from observed ICL (5'-^{32}P-t-**155**, NC = 1.45 ± 0.22), indicating greater distortion in the former (Figure 53A and Table 5).

Finally, comparing reactivity at T_{50} (four nucleotides away from ICL) between 5'-^{32}P-t-**155** and 5'-^{32}P-t-**156** using the same $KMnO_4$ reaction above (App. Figure 50) showed that T_{50} in 5'-^{32}P-t-**156** (unobserved ICL) was more reactive than that in 5'-^{32}P-t-**155** (observed ICL) (Figure 53B and Table 5). The higher NC value (reference nucleotide G_{54}) for 5'-^{32}P-t-**156** (unobserved ICL, NC = 2.14 ± 0.27) was intriguing since enhanced reactivity at opposing A_{11} was not observed. This could either be due to different extent of solvent exposures for the two strands or from possible low selectivity of the DEPC reaction. DEPC reaction carried out on A_{51} (five nucleotides away from ICL) did not detect any significant difference in the reactivity between 5'-^{32}P-t-**155** and 5'-^{32}P-t-**156**, implying reduced distortion (not shown).

3.3.1.1.5. Investigation of hyperreactive sites on the 5'-side of the top strand (abasic site analogue containing strand)



This region was not covalently attached to the cross-link but was hydrogen bonded to the 3'-side of the bottom strand (cross-linked A containing strand). We

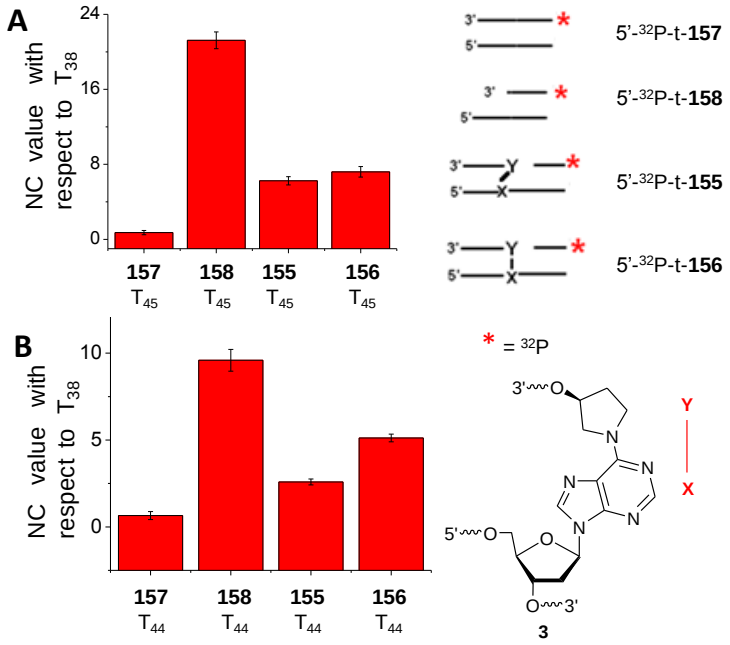


Figure 54. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 5'-^{32}P-t-**155** (observed ICL), 5'-^{32}P-t-**156** (unobserved ICL), 5'-^{32}P-t-**157** (control duplex), and 5'-^{32}P-t-**158** (control to assess the effect of strand break) for $KMnO_4$ reaction. (A) Comparison of ratio of $KMnO_4$ cleavage intensity of T_{45} from 5'-^{32}P-t-**155**, 5'-^{32}P-t-**156**, 5'-^{32}P-t-**157** and 5'-^{32}P-t-**158** with respect to T_{38}. (B) Comparison of ratio of $KMnO_4$ cleavage intensity of T_{44} from 5'-^{32}P-t-**155**, 5'-^{32}P-t-**156**, 5'-^{32}P-t-**157** and 5'-^{32}P-t-**158** with respect to T_{38}.

expected the 3'-side of the nick to have minimal structural destabilization, but possibly some denaturation (and increased reactivity) caused by the adjacent strand break. Since this destabilization was anticipated to be confined to the region closest to the nick, which consisting primarily of Ts, **155** (observed ICL), **156** (unobserved ICL), and **157** (control duplex) ^{32}P-labeled at the 5'-terminus of the top strand (5'-^{32}P-t-**155**, 5'-^{32}P-t-**156**, and 5'-^{32}P-t-**157**) were treated with $KMnO_4$. Substrate **158** ^{32}P-labeled at the 5'-side of the shorter or "top" strand (5'-^{32}P-t-**158**) was used as a control to gauge any effect of the strand break. $KMnO_4$ treatment detected hyperreactivity at T_{45} and T_{44} for 5'-^{32}P-t-**155**, 5'-^{32}P-t-**156**, and 5'-^{32}P-t-**158** (App. Figure 53). Partial denaturation at T_{45} and T_{44} caused

by the proximity of a single stranded region was believed to be responsible for the significant hyperreactivity for 5'-^{32}P-t-**158.** In addition, the NC values [$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$, reference nucleotide T_{38}] for T_{45} and T_{44} for

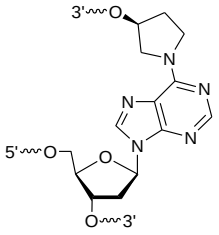
Table 6. Chemical probing experiments on DOB ICL analogue 3 for 5'-side of abasic side analog containing strand ("top" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
T_{45} of **157**	0.72 ± 0.23	T_{38} ($KMnO_4$)
T_{45} of **158**	21.23 ± 0.89	
T_{45} of **155**	6.25 ± 0.44	
T_{45} of **156**	7.20 ± 0.56	
T_{44} of **157**	0.65 ± 0.23	T_{38} ($KMnO_4$)
T_{44} of **158**	9.59 ± 0.63	
T_{44} of **155**	2.59 ± 0.17	
T_{44} of **156**	5.12 ± 0.22	

5'-^{32}P-t-**156** (NC = 7.20 ± 0.56 and 5.12 ± 0.22, respectively) and 5'-^{32}P-t-**155** (NC = 6.25 ± 0.44 and 2.59 ± 0.17, respectively) were lower than that for 5'-^{32}P-t-**158** (NC = 21.23 ± 0.89 and 9.59 ±0.63, respectively) but more than the duplex 5'-^{32}P-t-**156** (NC = 0.72 ± 0.23 and 0.65 ± 0.23, respectively) (Table 6 and Figure 54). This revealed the effect of the strand break in addition to possible hyperreactivity from ICL induced distortion, although the extent of the contribution from each is difficult to ascertain (Figure 54A and B, and Table 6). Importantly, the NC values for both T_{45} and T_{44} were higher for 5'-^{32}P-t-**156** (unobserved ICL) than 5'-^{32}P-t-**155** (observed ICL) (Figure 54 and App. Table 5). Nucleotides further from the ICL than the T_{44} were not probed.

3.3.1.1.6. Summary

Chemical probing revealed that nucleotides in the cross-links **155** and **156** are more reactive than the un-cross-linked duplex, indicating the deformity caused by ICL formation. Among the chemical probing reagents used in this study, the hydroxylamine reaction was not very reproducible, and was associated with a large statistical error. While DEPC and DMS reacted with the A and G in canonical base pairs, they were significantly more sensitive towards distorted nucleotides adjacent to the DOB ICL analogue. In contrast, $KMnO_4$ detected distorted nucleotides more efficiently. Its very low reactivity with the canonical Ts of a duplex indicated its effectiveness for detecting distortion. Hydroxyl radical cleavage, on the other hand, efficiently detected solvent exposure of cross-linked nucleotides. The results of the chemical probing experiments are summarized as a histogram (Figure 55) and cartoon (Figure 56). For the 5'-side of the cross-linked A containing strand, the G_{13} on **156** (unobserved ICL) was more distorted



3

(has greater NC value) than **155** (unobserved ICL) (Figure 55 and 56). On the same side of ICL **3**, the average NC values were higher for unobserved ICL (**156**) over observed ICL (**155**) for nucleotides A_{11} and C_{12}. However, the associated errors are large for both of them. For the 3'-side of the cross-linked A containing strand, the structural distortion was more prominent at A_{16} and A_{17} for unobserved ICL (**156**) than observed ICL (**155**)

(Figure 55 and 56). On the 3'-side of the abasic site analogue containing strand (top strand), the hyperreactivity of nucleotides towards chemical agents was detected up to

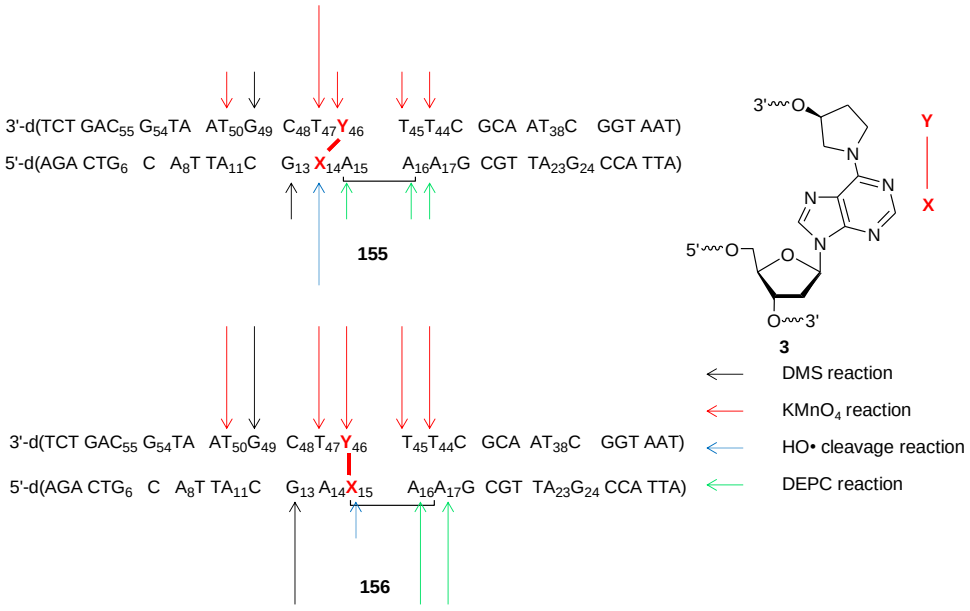


Figure 55. Histogram describing qualitative comparison of chemical probing on oligonucleotides **155** (observed ICL) and **156** (unobserved ICL) containing DOB ICL analogue **3**. The length of the arrows represents the reactivity difference for a particular nucleotide (towards a specific chemical agent). The differences in length for two different nucleotides are not comparable.

T_{50}. The nucleotides in this region were generally more reactive for **156** (unobserved ICL) than **155** (observed ICL), including the position of cross-linking (Y_{46}) itself (Figure 55 and 56). The only exception was the higher reactivity of T_{47} in **155** (observed ICL), which was probably due to partial or complete loss of H-bonding from the opposing A_{14}. The greater distortions for nucleotides in **156** (unobserved ICL) continued through C_{48} (albeit with high standard deviation) to G_{49} until T_{50}, beyond which the distortion could not be distinguished from that in **155** (observed ICL). Similarly, T_{45} and T_{44} in the un-cross-linked region 5' to the abasic site analogue were more solvent exposed for unobserved ICL **156** (Figure 55 and 56). The statistical significance of the difference in NC value between **155** and **156** was verified using Student's t test.

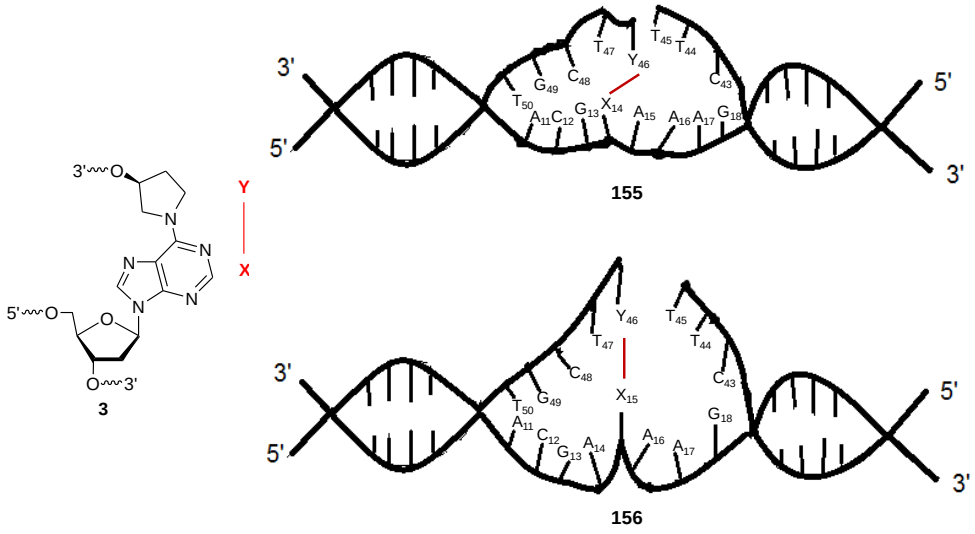


Figure 56. Cartoon depicting effect of DOB cross-link analogue **3** on **155** (observed ICL) and **156** (unobserved ICL).

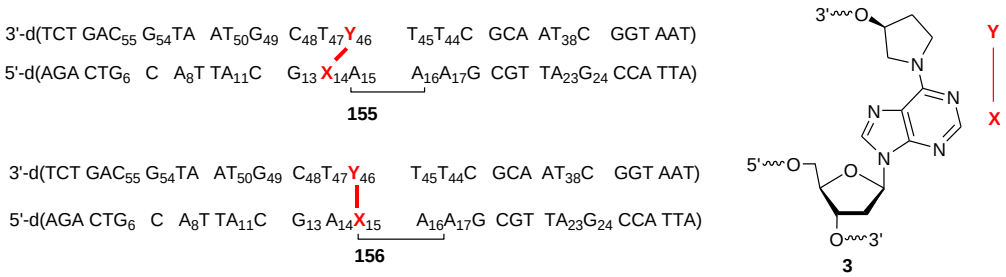
Hydroxyl radical cleavage showed that the solvent exposure of the cross-linked nucleotide X_{14} (for **155)** or X_{15} (for **156**) in the bottom strand was higher in **155** than **156**. Interestingly, this is opposite to what was found for the top strand, where the reactivity of Y_{46} in **156** was significantly higher than that in **155** (Figure 55 and 56). We theorized that the cross-link pushed the X_{15} in **156** on the bottom strand into the helix backbone, in turn protruding out a highly exposed pyrrolidinium ring on top strand (Figure 56). We wondered whether unequal solvent exposure detected for the cross-linked A and abasic site analogue (characterized by hydroxyl radical cleavage) could originate from their different stabilization by helix backbone. The cross-linked A in the bottom strand can be stabilized by base stacking or partial H-bonding (Figure 56). In contrast, there is no possible way to stabilize the pyrrolidinium ring at the top strand adjacent to a nick, resulting in higher solvent exposure.

A greater number of hyperreactive nucleotides were detected for **155** and **156** in the abasic site analogue containing top strand (distortion detected until T_{50}, fifth

nucleotide from cross-link position, Figure 55) than the cross-linked A containing strand (distortion detected until dG_{13}, third nucleotide from cross-link). This is in accordance with the unequal degree of distortion found in the strands connected by an asymmetric ICL.[88] Whether the asymmetric distortion seen for **3** is genuine or is a result of unequal reproducibility and sensitivity of the chemical probing agents is hard to determine from the current set of experiments. If this observation is real, this might be due to the translation of the distortion of the cross-linked pyrrolidinium ring into the backbone (Figure 55 and 56). It likely produced a relatively longer stretch of distorted nucleotides in the top strand. An observation supporting this theory was the presence of nucleotides with greater NC value for the top strand of **156** than **155**, induced by a more distorted pyrrolidine ring. On the other hand, a less distorted A in the bottom strand of **155** and **156** was unable to influence that many nucleotides, thereby producing only 1 – 2 nucleotides with higher NC values for the latter (Figure 55 and 56).

3.3.1.2. Melting temperature analysis of cross-linked DNA containing DOB ICL analogue 3

3.3.1.2.1. Substrate design for melting temperature studies



Despite chemical probing being a reliable tool for detecting distortion, we wanted to validate our findings about instability in cross-linking with opposing nucleotides using a more quantitative method, the oligonucleotide melting temperature. We anticipated the cross-linked region of the substrates containing DOB ICL analogue **3** would behave like

a hairpin, increasing the melting temperature.[149] Oligonucleotides **85** and **86** containing **3** were prepared based on a theoretical T_m calculation using hypothetical hairpin substrate **159** (Figure 57). Oligonucleotides **85** and **86** contained the "observed" (cross-link with A opposite 3'-adjacent T) and "unobserved" (cross-linking with opposing A) types of ICL and were based on chemical probing substrates **155** and **156**, respectively. The stem of

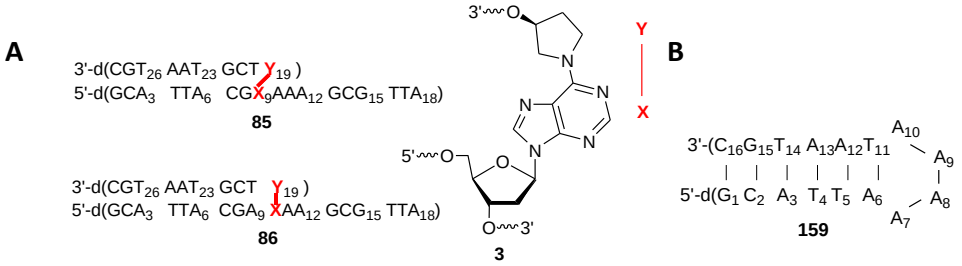


Figure 57. Cross-links containing DOB ICL analogue **3** and their hypothetical hairpin model. (A) Cross-linked oligonucleotides made using solid phase synthesis. (B) Hypothetical hairpin used to estimate T_m of cross-link substrates **86** and **89**.

159 contained of the G_6 – A_{11} sequence in **155** and **156** that did not show any hyperreactivity during the chemical probing studies. The A_4 loop mimicked the region displaying highest reactivity against chemical reagents, 5'-$C_{12}G_{13}X_{14}Y_{46}T_{47}C_{48}G_{49}$-3' in **155** and 5'-$C_{12}G_{13}A_{14}X_{15}Y_{46}T_{47}C_{48}G_{49}$-3' in **156**. We anticipated that the vicinity of the A_6 – T_{11} pair in **159** with the loop would mimic the relatively low-level distortion detected in the A_{11} – T_{50} pair in **155** and **156**. A calculation using an online T_m calculator (https://www.idtdna.com/calc/analyzer) predicted ~60 °C melting temperature for **159** (DNA concentration 1 μM, 100 mM Na^+, 10 mM Mg^{2+}). Substrates **85** ("observed" type of ICL) and **86** ("unobserved" type of ICL) were converted into **160** and **161** with fully complementary arms (Figure 58). The un-cross-linked region in **160** (full length observed ICL) and **161** (full length unobserved ICL) was estimated to have a 32.5 °C T_m based on the online T_m calculator (DNA concentration 1 μM, pH 7.0, 100 mM Na^+, 10 mM Mg^{2+}).

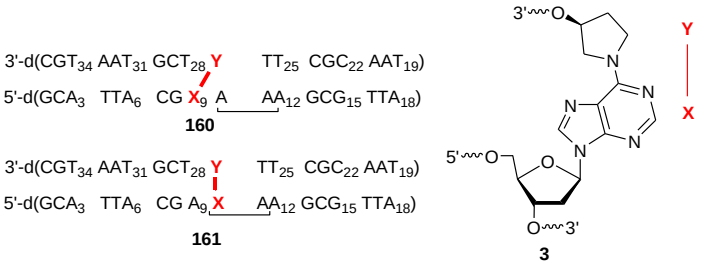
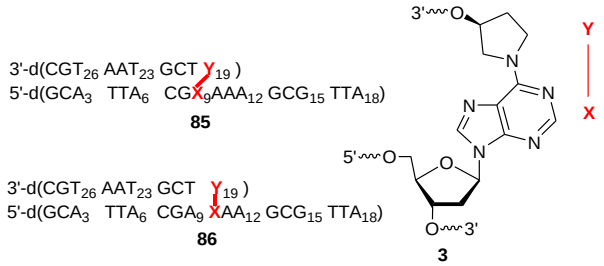


Figure 58. Cross-linked oligonucleotides containing DOB ICL analogue **3** used for melting temperature study.

3.3.1.2.2. Calculation of UV-extinction coefficients of melting temperature substrates

The extinction coefficient of **85** (observed ICL) and **86** (unobserved ICL) at 260 nm were calculated based on the nearest-neighbor model and a list of corrected extinction coefficients available online (http://www.owczarzy.net/extinctionDNA.htm).[156–158] The list and methodologies from the nearest-neighbor model are combined in a standard extinction coefficient calculator (http://biophysics.idtdna.com/UVSpectrum.html). To simplify the calculation, the hairpin like regions of **85** and **86** were assumed to remain base paired even in the absence of salt. Based on this premise, nucleotides $G_1 – X_9$ in **85** were considered as a duplex, with X having the chromophore of A. Since the cross-linked abasic site (Y_{19}) lacked a chromophore, it did not contribute towards the extinction coefficient. Similarly, we envisioned nucleotides $G_1 – A_9$ in **86** to form a duplex. Both duplexes are estimated to have Ɛ = 144544



$L \cdot mol^{-1} \cdot cm^{-1}$ based on the standard extinction coefficient calculator (http://biophysics.idtdna.com/UVSpectrum.html). The next nucleotides, $A_{10} - A_{18}$ in **85** and $X_{10} - A_{18}$ in **86** were treated as a continuous single stranded chain of residues with calculated ε = 97000 $L \cdot mol^{-1} \cdot cm^{-1}$. According to the nearest neighbor method, the extinction coefficients of **85** and **86** would approximately be the sum of the single stranded and duplex portions, 241544 $L \cdot mol^{-1} \cdot cm^{-1}$. We estimate an error of less than 5% in this calculation due to the direct addition of extinction coefficients of the single stranded and duplex regions.

3.3.1.2.3. Melting temperature study of oligonucleotides containing DOB ICL analogue 3

The melting temperatures of **160** (observed ICL) and **161** (unobserved ICL) show that each has two independent melting transitions (Figure 59A). The melting occurring at higher temperature was in the range of 60-70°C, while the other was around 30°C. To identify the structural part of the mimics that contributed to different range of T_m values, we did melting temperature studies on **85** ("observed" type of ICL) and **86** ("unobserved" type of ICL), the cross-links without a duplex un-cross-linked region (Figure 59B). Cross-links **85** and **86** had T_m values in the range 60 - 70°C (Figure 59B and Table 7), further implying that the lower temperature T_m values were from the denaturation of un-cross-linked regions of the observed (**160**) and unobserved (**161**) ICL. The higher T_m values were close to the predicted values of model hairpin substrate **159** and online T_m calculator, confirming the hairpin like behavior of two complementary strands connected through a cross-link. Interestingly, T_m values of **160** and **161** in the un-cross-linked region were lower than that compared to the predicted T_m (32.8°C) under the

same salt and oligonucleotide concentrations (Figure 59C and Table 7). This can be an indirect corroboration of the deformities encountered in this region during the chemical probing studies.

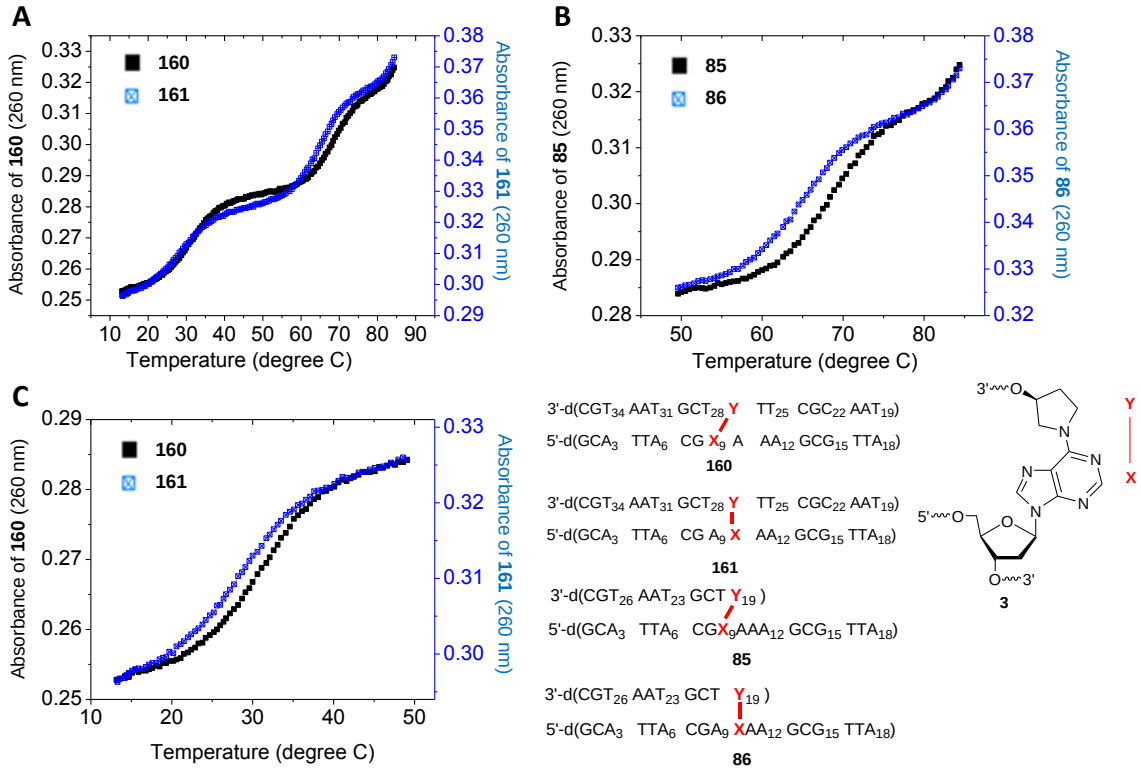


Figure 59. Thermal denaturation analysis of oligonucleotides containing DOB ICL analogue **3**. (A) Absorbance (260 nm) vs temperature profile of cross-linked substrates **160** (observed ICL) and **161** (unobserved ICL) [DNA concentration 1 µM, PIPES (pH 7.0) 10 mM, Mg^{2+} 10 mM, Na^{+} 100 mM]. (B) Absorbance (260 nm) vs temperature profile of cross-linked substrates **85** (containing observed type of cross-link) and **86** (containing observed type of cross-link) [DNA concentration 1 µM, PIPES (pH 7.0) 10 mM, Mg^{2+} 10 mM, Na^{+} 100 mM]. (C) Absorbance (260 nm) vs temperature profile of cross-linked substrates **160** and **161** in the temperature range 10 – 50 °C [DNA concentration 1 µM, PIPES (pH 7.0) 10 mM, Mg^{2+} 10 mM, Na^{+} 100 mM]. (D) Comparison of higher T_m values of **160** (observed ICL) and **161** (unobserved ICL). (E) Comparison of lower T_m values of **160** and **161**.

Oligonucleotide **160** (observed ICL) had higher T_m values in both regions compared to those of **161** (unobserved ICL) (Figure 59, Table 7). The un-cross-linked region in the observed ICL (**160**) melted approximately 2 °C higher than unobserved ICL (**161**) (Figure 59E, Table 7). For the higher T_m region, the melting temperature differed slightly for substrates with and without a duplex un-cross-linked region. While substrates

85 (observed ICL) and **86** (unobserved ICL) had T_m values 70.15 ± 0.40 °C and 66.28 ± 0.35 °C, the T_m values of full length ICLs **160** and **161** were 69.76 ± 0.83 °C and 65.32 ± 0.03 °C in the temperature range, respectively. Since the un-cross-linked region (T_m ~30 °C) is dehybridized at the higher temperature, the difference was considered an artifact. Nonetheless, it can be equivocally asserted that cross-link with an A opposite 3'-adjacent nucleotide ("observed", present in **85** and **160**) as observed in the native DOB ICL is thermodynamically more stable than an ICL with opposing A ("unobserved", present in **86** and **161**). This observation is in complete accord with the chemical probing studies, where nucleotides in **156** (consisting of cross-link with opposing A) generally displayed more sensitivity towards chemical probing agents. Together, the melting temperature study and chemical probing analysis supported our hypothesis that cross-link with opposing A is associated with greater deformity in the helix.

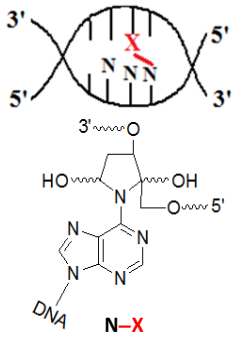
Table 7. T_m of cross-linked oligonucleotides 85, 86, 160 and 161 [DNA concentration 1 µM, PIPES (pH 7.0) 10 mM, Mg^{2+} 10 mM, Na^+ 100 mM]

Substrate	Tm (degree C)
Observed ICL **(160)**	69.76 ± 0.83 and 30.95 ± 0.26
Unobserved ICL **161**	65.32 ± 0.03 and 28.12 ± 0.55
85 (ICL with observed type of cross-link)	70.15 ± 0.40
86 (ICL with unobserved type of cross-link)	66.28 ± 0.35

3.3.2. Structural probing of C4-AP ICL mimics

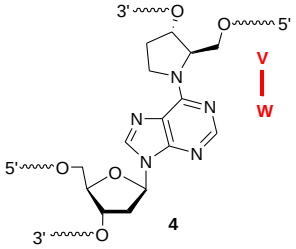
3.3.2.1. Chemical probing of cross-linked DNA containing C4-AP ICL analogue 4

3.3.2.1.1. Chemical probing reactions and substrate design



C4-AP kinetic ICL (**2**)

C4'-oxidized abasic site (C4-AP) forms the kinetic interstrand cross-link (**2**) with an A opposite 3'-adjacent nucleotide in a site selective manner (Scheme 10). While the typical yield of the kinetic cross-link with A opposite 3'-adjacent nucleotide varies from 16 – 37%, no cross-linking is detected with an opposing A.[8] We hypothesized that the inherent structural stability of the respective adducts could be a deciding factor behind the site selectivity. To probe any possible structural distortions using the chemical probing methodology and melting temperature analysis, we used the synthesized stable



analogue of C4-AP ICL (**4**). Diethylpyrocarbonate (DEPC), $KMnO_4$, hydroxyl radical cleavage and dimethyl sulfate (DMS) were the most useful chemical probing techniques used to analyze the DOB ICL analogue (**3**) in terms of reproducibility. Consequently, the

flanking sequence of the chemical probing substrates containing **4** were comprised predominantly of As and Ts. Although DMS treatment employed for probing G was useful in detecting distortion, the presence of G in the sequence required that we include C (to be probed by inefficient hydroxylamine) in the opposing strand. Since the hydroxylamine reaction was unreliable in our hands, the flanking sequence of **4** was deprived of Gs.

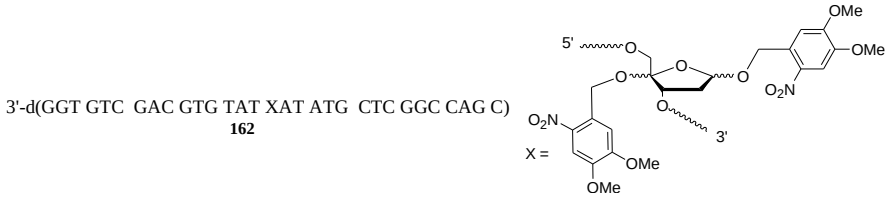


Figure 60. Single stranded and duplex oligonucleotides used to determine the site-selectivity of C4-AP cross-link formation.

To test whether the flanking sequence comprised of As and Ts would lead to the previously observed site-selectivity for C4-AP ICL formation, we synthesized oligonucleotide **162** containing a 4,5-dimethoxy-2-nitrobenzyl protected C4-AP precursor (Figure 60).[8] After hybridization of **162** with the complementary strand, photoirradiation (350 nm) of the resulting duplex (**163**) produced C4-AP.[8] Incubation of the irradiated duplex labeled at 37 °C and analysis of the aliquot thereof by denaturing PAGE indicated cross-link formation detected as two slower moving bands (kinetic and thermodynamic ICLs). Purification of the slower moving band (high molecular weight or kinetic C4-AP ICL) and its treatment with hydroxyl radical indicated ICL formation with A_{15}, the nucleotide opposite the 3'-adjacent T (App. Figure 54). Having proven the site-selectivity of C4-AP ICL formation in the sequence, we constructed **121** and **164** using synthetic

methodologies developed earlier (Figure 61). While **121** had the "observed" cross-link with A opposite the 3'-adjacent T, **164** was comprised of the "unobserved" cross-link

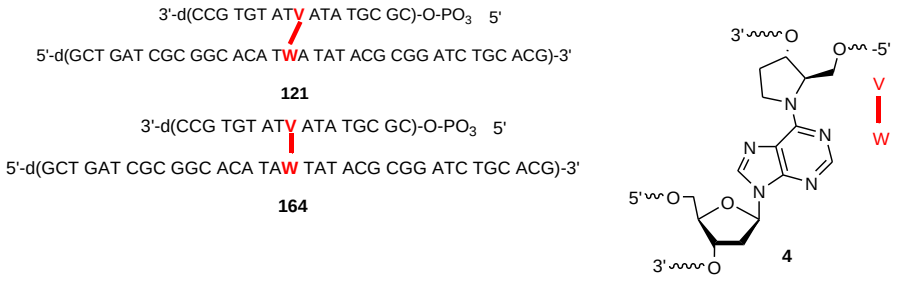
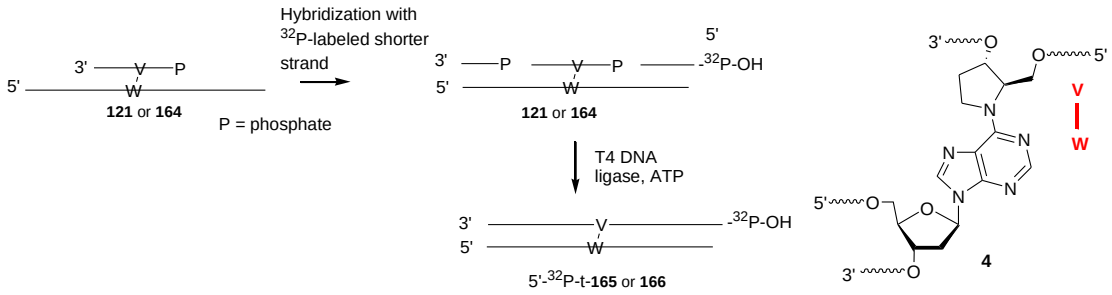


Figure 61. Oligonucleotides used for making the chemical probing substrates on 4 and cartoons showing overhang and recessed end. (A) Sequences of oligonucleotides containing the "observed" (**121**) and "unobserved" (**164**) C4-AP ICL analogue **4**. (B) Cartoon showing an example of a 5'-overhang. (C) Cartoon showing an example of a 5'-recessed end.

with the opposing A. Using ligation, substrates **121** and **164** were converted into **165** and **166** containing full length complementary arms (Figure 62). See Experimentals for details regarding the construction of **165** and **166**.. The ligation strategy also enabled independent ^{32}P-labeling of all three termini except the 5'-side of the cross-linking

Scheme 42. Conversion of 121 or 164 into 5'-^{32}P-t-165 or 166



A containing strand ("bottom" strand). Similar to the chemical probing of DOB ICL analogue **3**, the reactivity of an ICL-adjacent nucleotide of a ^{32}P-labeled oligonucleotide was represented using the "normalized cleavage" or NC value $[\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}]$ with reference to a distant nucleotide. In addition, we used duplex **167** as a control substrate to identify hyperreactivity induced by

local sequence (Figure 62). The cross-linked A and abasic site analogue in **165** were replaced by A and T in **167**. The difference in reactivity of nucleotides in observed ICL (**165**) and unobserved ICL (**166**) were compared using Student's t test (App. Table 5 – 8).

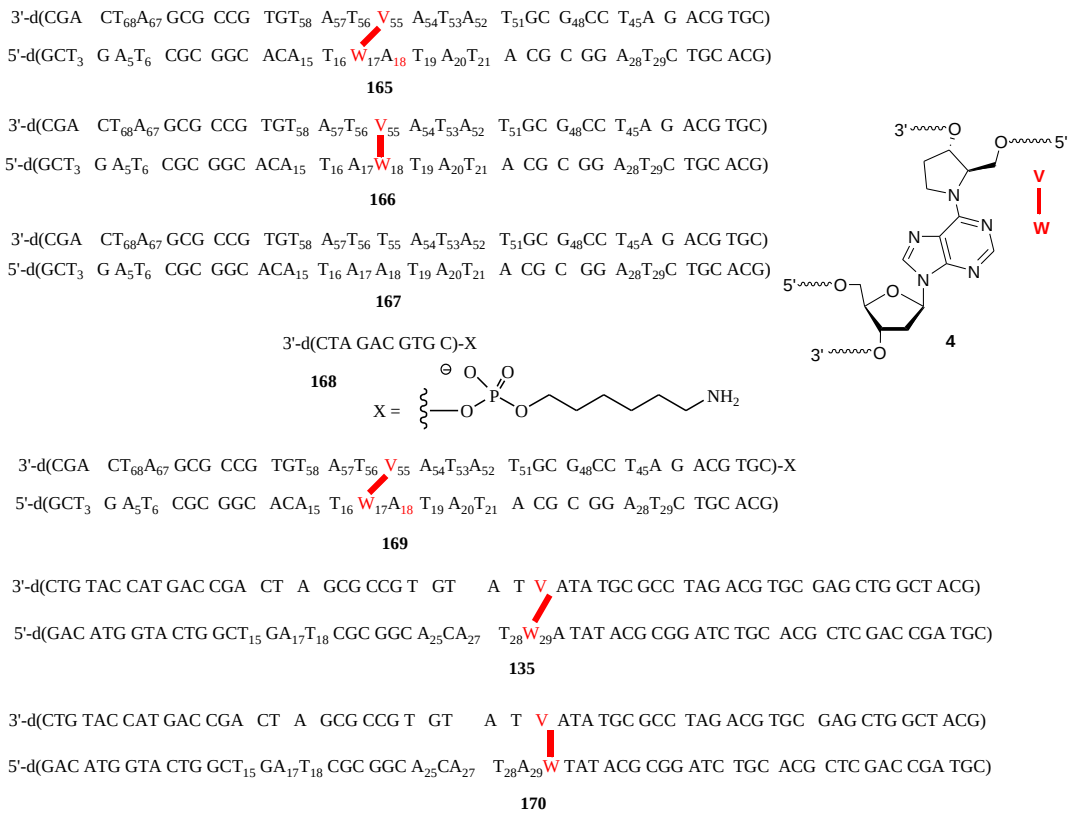
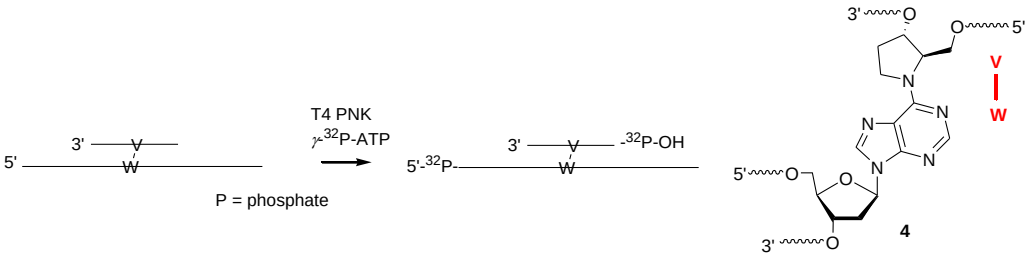


Figure 62. Oligonucleotides used for chemical probing of C4-AP ICL analogue **4**.

In order to probe the nucleotides on 5'-side of the bottom strand (cross-linked A containing strand, nucleotides G_1 – A_{18} / W_{18}, Figure 62) or on the 5'-side of the top strand (the abasic site analogue containing strand, nucleotides V_{55} – G_{37}, Figure 62) of **165** and **166**, the respective ends needed to be independently ^{32}P-labeled. While the 5'-termini of the top strand can be easily labeled by hybridization with a ^{32}P-labeled small strand with **121** or **164** followed by enzymatic ligation (Scheme 41, see Experimentals for further details), the 5'-^{32}P-labeling of the bottom strand was trickier. The 5'-^{32}P-labeling of the bottom strand (an overhang, Figure 61) by incubation with T4 PNK and γ-^{32}P-ATP

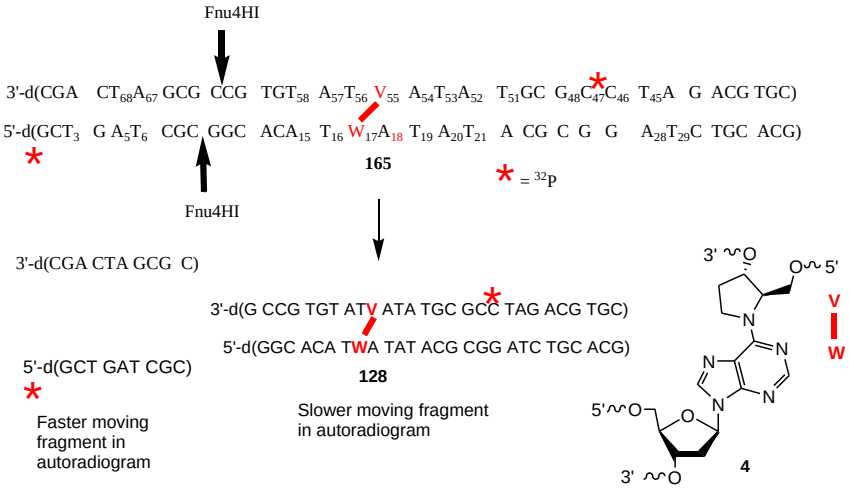
can cause unwanted 5'-^{32}P-labeling of the top strand (a recessed end, Figure 61) as well (Scheme 42). To avoid this, substrates **121** and **164** were chemically attached with

Scheme 43. Unwanted 5'-^{32}P-labeling of a substrate containing two 5'-terminus



phosphate at the 5'-terminus of abasic site analogue containing strand ("t" or "top" strand) during the solid phase synthesis (Figure 61). A phosphate present at the 5'-recessed end is 15 – 20% less efficiently exchanged than that on an overhang or single strand.[159] In addition, the exchange reaction by T4 PNK becomes 86th times slower at 0 °C, while the phosphorylation at 5'-OH remains significantly faster.[160] Therefore, we

Scheme 44. Fnu4HI treatment of internally labeled 5'-^{32}P-b-165



theorized that treatment of **121** and **164** with T4 PNK and γ-^{32}P-ATP at 0 °C would significantly slow the exchange of the 5'-phosphate on the top strand with ^{32}P-phosphate. Under the same conditions, the ^{32}P-labeling of 5'-terminus of the cross-linked A

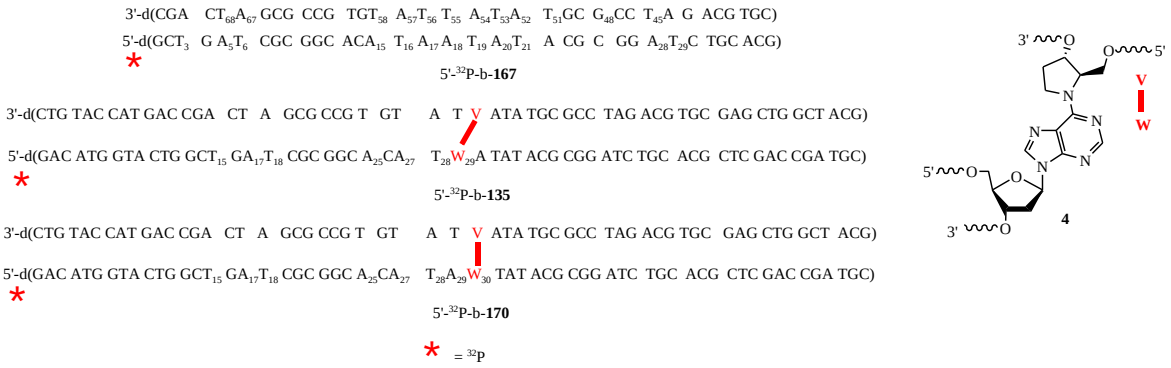
containing strand ("b" or "bottom" strand) would be considerably faster. The theory was tested to achieve selective ^{32}P-labeling of the bottom strand of **121**. After the incubation, the T4 PNK was inactivated (heating at 65 °C) and removed from solution (by phenol-chloroform-isopropanol wash), while **121** was subjected to ligation to make **165.** See Experimentals for details regarding the construction of **165**. The phosphate exchange would lead to internal labeling at top strand and was determined by treatment of the 5'-^{32}P-b-**165** with restriction enzyme Fnu4HI. Absence of internal labeling would completely consume 5'-^{32}P-b-**165** and would only produce a 9 nucleotide long product in denaturing PAGE (Scheme 44). If internal labeling occurs, a slower migrating band underneath the starting material would be seen in the gel (Scheme 44 and App. Figure 55). Analysis of the aliquot from Fnu4HI treatment of 5'-^{32}P-b-**165** revealed the presence of up to 32% of internal labeling originating from the phosphate exchange (App. Figure 56A). In an alternative strategy, **121** was ligated with a 5'-C6-amino modified oligonucleotide **168** to produce **169** prior to incubation with T4 PNK and γ-^{32}P-ATP. The presence of the 5'-C6-alkyl amino modification prevents any 5'-^{32}P-labeling on the top strand, and promotes selective labeling of the bottom strand. Although unlikely, any occurrence of ^{32}P-labeling on the top strand was probed using treatment with Fnu4HI, which would produce a slow moving band (App. Figure 55). Treatment of 5'-^{32}P-b-**169** (the bottom strand is referred to as "b") with Fnu4HI revealed ~11% ^{32}P-labeling at top strand (App. Figure 56B). We theorized that the undesired labeling at the top strand resulted from an impurity lacking the C6-amino modification present in **168**, which probably fell off during purification. As a result, a small fraction of **169** prepared from ligation of **168** was missing the C6-alkyl amino modification and was subsequently ^{32}P-

labeled at the top strand. Since even a small amount of incorrectly located ^{32}P-labeling could jeopardize the sensitive chemical probing experiments, the selective 5'-labeling strategies were bypassed. However, the selective ^{32}P-labeling strategy based on C6-alkyl amino modification can succeed if **169** is carefully prepared without any impurity.

To overcome this problem, ICLs **135** and **170** were synthesized specifically for probing the nucleotides on the 5'-side of the bottom strand (Figure 62). We intended to avoid using **135** and **170** for chemical probing for two fold reasons. We correlated their longer length with possibly employment of longer reaction time and higher reagent concentration. In addition, we envisioned that four ligations should be needed to construct substrates **135** and **170** would yield lower final products in contrast with two ligations required for **165** and **166**. Both our suspicions were proved wrong with similar final isolation yield (4 – 10%) and same chemical agents concentrations used. Ligation strategies developed earlier for synthesizing nucleotide excision repair substrates were employed to make **135** and **170**. It involved phosphorylation of **121** and **164** using cold ATP, thereby eliminating any chance of ^{32}P-phosphate exchange on the top strand and the possibility of internal labeling in full length **170** and **171**. The integrity of 5'-^{32}P-b-**171** was verified by treatment with restriction enzymes RsaI and Taq$^{\alpha}$I (App. Figure 55 and 57). The difference in hyperreactivity between **165** and **167** (alternatively, between **170** and **171**) were analyzed with Student's t test to verify their statistical significance.

3.3.1.1.2. Reactivity of nucleotides on the 5'-side of the cross-linked A containing strand

In order to detect hyperreactive nucleotides at the 5'-side of the cross-linked A containing strand, treatment of 5'-^{32}P-b-**167** (control duplex), 5'-^{32}P-b-**135** (observed



ICL), and 5'-^{32}P-b-**171** (observed ICL) (The A containing strand was referred to as the "b" or "bottom" strand, while the abasic site analogue containing strand was denoted as "t" or "top".) required greater concentrations of chemical probing agents than their DOB counterparts. Compared to 0.3 mM $KMnO_4$ (final concentration, 10 min, room

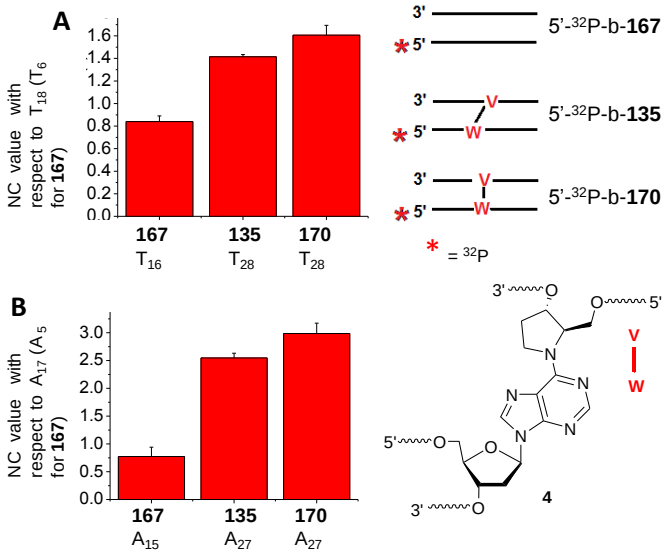


Figure 63. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 5'-^{32}P-b-**167** (control duplex), 5'-^{32}P-b-**135** (observed ICL), and 5'-^{32}P-b-**170** (unobserved ICL) for DEPC and $KMnO_4$ reaction. (A) NC value for $KMnO_4$ reaction of T_{28} (T_{16} for 5'-^{32}P-b-**167**) from 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170** with reference to T_{18} (T_6 for 5'-^{32}P-b-**167**). (B) NC value for DEPC reaction of A_{27} (A_{15} for 5'-^{32}P-b-**167**) from 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170** with reference to A_{17} (A_5 for 5'-^{32}P-b-**167**).

temperature) and 1.05 M DEPC (final concentration, 12 min, room temperature) treatment for DOB substrates, the C4-AP substrates irrespective of the location of ^{32}P-labeling required 0.6 mM $KMnO_4$ (final concentration, 10 min, room temperature) and 1.75 M DEPC (final concentration, 12 min, room temperature) to achieve similar band

intensity on the denaturing PAGE. This could be attributed to less distortion of the C4-AP substrates than the DOB ones. In addition, the C4-AP substrate concentration ranged from 5 – 10 nM, which was smaller than the DOB reaction concentration (30 – 60 nM). Assuming single hit kinetics (second order reaction), an equal amount of product generation (detected by percentage cleavage) would require a higher concentration of reagent for C4-AP, which is indeed the case.

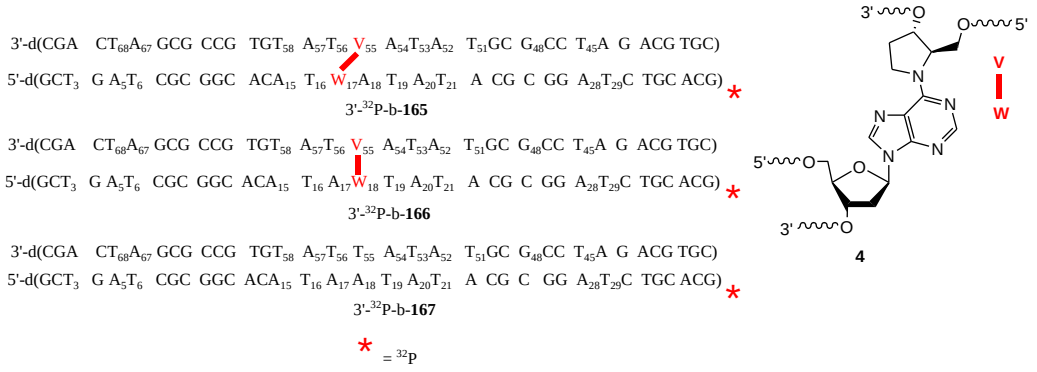
Table 8. Chemical probing experiments on C4-AP ICL analogue 4 for 5'-side of A containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**135**), unobserved ICL (**170**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
T_{16} of **167**	0.84 ± 0.05	T_{18} (T_6 for **167**) ($KMnO_4$)
T_{28} of **135**	1.42 ± 0.02	
T_{28} of **170**	1.60 ± 0.09	
A_{15} of **167**	0.77 ± 0.17	A_{17} (A_5 for **167**) (DEPC)
A_{27} of **135**	2.43 ± 0.03	
A_{27} of **170**	2.54 ± 0.02	

Unlike the oligonucleotides containing DOB ICL analogue **3**, the cross-linked A in 5'-^{32}P-b-**135** and 5'-^{32}P-b-**170** (W_{29} and W_{30}, respectively) was cleaved upon subjection to the A-G sequencing reaction (formic acid, room temperature) (Lanes 6 and 12, App. Figure 58). The difference in reactivity could not be rationalized in any way. The reaction of 5'-^{32}P-b-**135** (observed ICL) and 5'-^{32}P-b-**171** (unobserved ICL) with $KMnO_4$ detected hyperreactivity at T_{28} (nucleotide adjacent to ICL) (App. Figure 58). Unlike the Ts in control duplex (**157**) used for DOB substrates, the Ts present in its C4-AP counterpart **167** demonstrated significant reactivity, presumably due to the higher $KMnO_4$ concentration employed. The NC value [$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$,

reference nucleotide T_{18} (T_6 for **167**)] detected greater hyperreactivity for 5'-^{32}P-b-**171** (unobserved ICL, NC = 1.60 ± 0.09) (Figure 63A and Table 8). Due to the presence of multiple As in the sequence bordering the cross-link in this region, the treatment of 5'-^{32}P-b-**135** (observed ICL) and 5'-^{32}P-b-**170** (unobserved ICL) with DEPC was more informative. Hyperreactivity was detected at A_{29} of 5'-^{32}P-b-**170** (unobserved ICL) indicating partial or complete disruption of H-bonding with opposing T (App. Figure 59). Less hyperreactivity was detected for A_{27} in both substrates from the respective denaturing PAGE and histograms (App. Figure 59D and E). The NC value [reference nucleotide A_{17} (A_5 for control duplex **167**)] indicated the greater distortion for A_{27} in 5'-^{32}P-b-**171** (unobserved ICL, NC = 2.54 ± 0.02) (Figure 63B and App. Table 6). Examination of the DEPC histograms of observed ICL (5'-^{32}P-b-**170**) and unobserved ICL (5'-^{32}P-b-**171**) revealed no hyperreactivity at any As 5' to it, indicating the disappearance of ICL induced distortion at A_{25} and beyond.

3.3.1.1.2. Reactivity of nucleotides on the 3'-side of the cross-linked A containing strand



In the region 3' to the cross-linked A (W_{17} in observed ICL **165** and W_{18} in unobserved ICL **166**), the reactivity of the nucleotides up to T_{21} were investigated using

their normalized cleavage or NC value [$\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}$]. Oligonucleotides 3'-^{32}P-b-**165** (observed ICL) and 3'-^{32}P-b-**166** (unobserved ICL, see Experimentals for their preparation) were subjected to hydroxyl radical cleavage reaction anticipating it would detect the solvent exposed sugar residues of the backbone, especially those of the cross-linked A. However, uniform cleavage intensity was

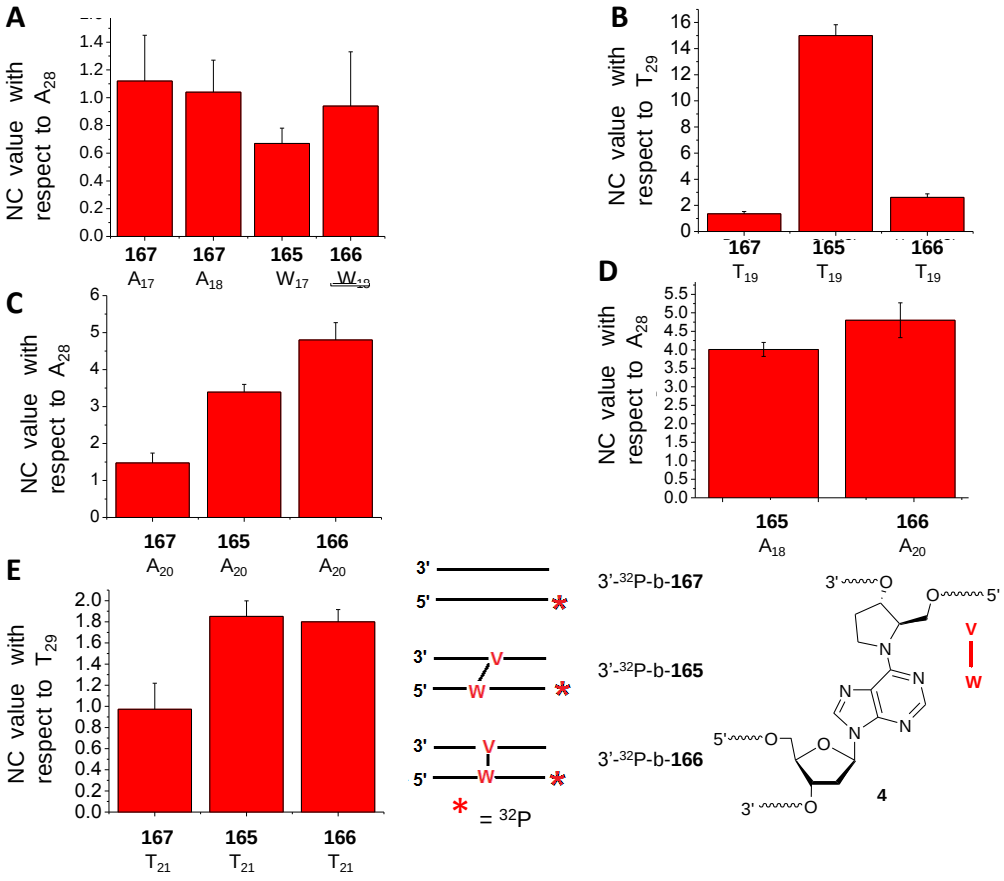


Figure 64. Comparison of cleavage of nucleotides (as a function of distance from ICL) (control duplex) in 3'-^{32}P-b-**167** (control duplex), 3'-^{32}P-b-**165** (observed ICL), and 3'-^{32}P-b-**166** (observed ICL) for hydroxyl radical cleavage, DEPC and $KMnO_4$ reaction. (A) NC value for hydroxyl radical cleavage reaction of A_{17} and A_{18} from 3'-^{32}P-b-**167** (W_{17} and W_{18} for 3'-^{32}P-b-**165** and 3'-^{32}P-b-**166**, respectively) with respect to A_{28}. (B) NC value for $KMnO_4$ reaction of T_{19} from 3'-^{32}P-b-**167**, 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** with respect to T_{29}. (C) NC value for DEPC reaction of A_{20} from 3'-^{32}P-b-**167**, 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** with respect to A_{28}. (D) NC value for DEPC reaction of A_{18} from 3'-^{32}P-b-**165** and A_{20} from 3'-^{32}P-b-**166** with reference to A_{28}. (E) NC value for $KMnO_4$ reaction of T_{21} from 3'-^{32}P-b-**167**, 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** with respect to T_{29}.

observed for for 3'-^{32}P-b-**165** and 3'-^{32}P-b-**166** in this region (App. Figure 60). The cross-linked A or the nucleotides near it (W_{17} and W_{18} for **165** and **166**, respectively) did not demonstrate any hyperreactivity compared to the distant ones. Instead, the cleavage pattern was similar to A_{17} or A_{18} in control substrate 3'-^{32}P-b-**167**, revealing little or no distortion in the cross-linked As. Corroborating this observation, the NC values (reference nucleotide A_{28}) for cross-linked As (W_{17} and W_{18} for **165** and **166**, respectively) were within the error limits of their counterparts in the duplex (A_{17} and A_{18} of **167**) (Figure 64A and Table 9). In contrast, the adjacent T_{19} was hyperreactive in 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166** (App. Figure 61). Interestingly, T_{19} in the observed ICL (3'-^{32}P-b-**165**, NC = 14.99 ± 0.84) was ~7 fold more reactive than that in the unobserved ICL **(**3'-^{32}P-b-**166**, NC = 2.61 ± 0.27) (App. Figure 61, Figure 64B and Table 9). This was opposite to the general trend of higher reactivity for nucleotides on 5'-side of the bottom strand of the unobserved ICL (**166**) (Figure 63). The detection of hyperreactivity at A_{18} during treatment of 3'-^{32}P-b-**165** was anticipated due to its position opposite cross-linked abasic site analogue (App. Figure 62). Hyperreactivity was detected in A_{20} as well in 3'-^{32}P-b-**165** (observed ICL) and 3'-^{32}P-b-**166** (unobserved ICL), with A_{20} in **166** being more sensitive towards DEPC (NC = 4.80 ± 0.47) (App. Figure 62, Figure 64C and Table 9). Interestingly, the reactivity of A_{20} (one nucleotide away from ICL) in the unobserved ICL (3'-^{32}P-b-**166**) was even higher than the ICL-adjacent A_{18} in observed ICL (3'-^{32}P-b-**165**, NC = 4.01 ± 0.19) (Figure 64D and Table 9). The histograms and NC value detected ICL-induced distortion present at the next nucleotides T_{21} of 3'-^{32}P-b-**165** (observed ICL, NC = 1.85 ± 0.14) and 3'-^{32}P-b-**166** (unobserved ICL, NC = 1.80 ± 0.11), where its

Table 9. Chemical probing experiments on C4-AP ICL analogue 4 for 3'-side of A containing strand ("bottom" strand)

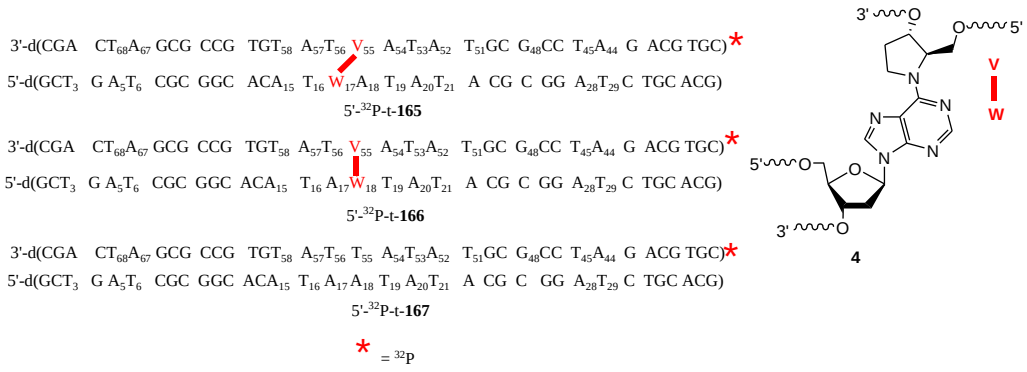
Nucleotide of interest in observed ICL (**165**), unobserved ICL (**166**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
A_{17} of **167**	1.12 ± 0.33	A_{28} (DEPC)
A_{18} of **167**	1.04 ± 0.23	
W_{17} of **165**	0.67 ± 0.11	
W_{18} of **166**	0.94 ± 0.39	
T_{19} of **167**	1.35 ± 0.17	T_{29} ($KMnO_4$)
T_{19} of **165**	14.99 ± 0.84	
T_{19} of **166**	2.61 ± 0.27	
A_{20} of **167**	1.47 ± 0.26	A_{28} (DEPC)
A_{20} of **165**	3.39 ± 0.21	
A_{20} of **166**	4.80 ± 0.47	
A_{18} of **167**	Not calculated	A_{28} (DEPC)
A_{20} of **167**	1.47 ± 0.26	
A_{18} of **165**	4.01 ± 0.19	
A_{20} of **166**	4.80 ± 0.47	
T_{21} of **167**	0.97 ± 0.24	T_{29} ($KMnO_4$)
T_{21} of **165**	1.85 ± 0.14	
T_{21} of **166**	1.80 ± 0.11	

reactivity was higher than that in the duplex **167** (App. Figure 61C and D, Figure 64E and Table 9). However, the NC values of 3'-^{32}P-b-**165** and 3'-^{32}P-b-**166** were within the error limit of each other (Figure 61E). This was presumably a sign of diminishing distortion at T_{21} and nucleotides 3' to it. Therefore, the nucleotides 3' to T_{21} were not probed.

3.3.1.1.3. Reactivity of nucleotides on the 5'-side of the cross-linked abasic site containing strand

In the region 5' to the cross-linked abasic site (V_{55}) containing strand ("t" or "top"), the reactivity of the nucleotides up to A_{52} were investigated using their normalized cleavage or NC value [$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]. Sensitivity towards DEPC was detected at A_{54} and A_{52} in substrates 5'-^{32}P-t-**165**

(observed ICL) and 5'-^{32}P-t-**166** (unobserved ICL) (see Experimentals for their preparation) (App. Figure 63 and 64). During the computation of the NC value, we



intended A_{44} to be the reference nucleotide. However, due to its location near the 5'-terminus, the band corresponding to A_{44} was often unclear at the lanes in denaturing PAGE. In other instances, merging of the lanes near the ^{32}P-labeling terminus due to salt complicated the cleavage intensity detection of A_{44}. Therefore, G_{48} was used as the reference nucleotide for computation of NC values for A_{54} and A_{52} in control duplex (5'-^{32}P-t-**167**), observed ICL (5'-^{32}P-t-**165**), and unobserved ICL (5'-^{32}P-t-**166**), despite G being a poor substrate of DEPC. The As in duplex control (5'-^{32}P-t-**167**) were significantly less reactive than their counterparts in the ICL substrates (5'-^{32}P-t-**165** and 5'-^{32}P-t-**166**) (App. Figure 63A). As a result of poor sensitivity of G towards DEPC and low reactivity of the As in the 5'-^{32}P-t-**167** (control duplex), the NC values of A_{54} and A_{52} for 5'-^{32}P-t-**165** (observed ICL) and 5'-^{32}P-t-**166** (unobserved ICL) were significantly higher than those for 5'-^{32}P-t-**167** (control duplex) (Figure 65A and Table 10). The NC values helped compare the sensitivity (distortion) of As in observed ICL (5'-^{32}P-t-**165**) and unobserved ICL (5'-^{32}P-t-**166**) to one another, high NC values should not be taken

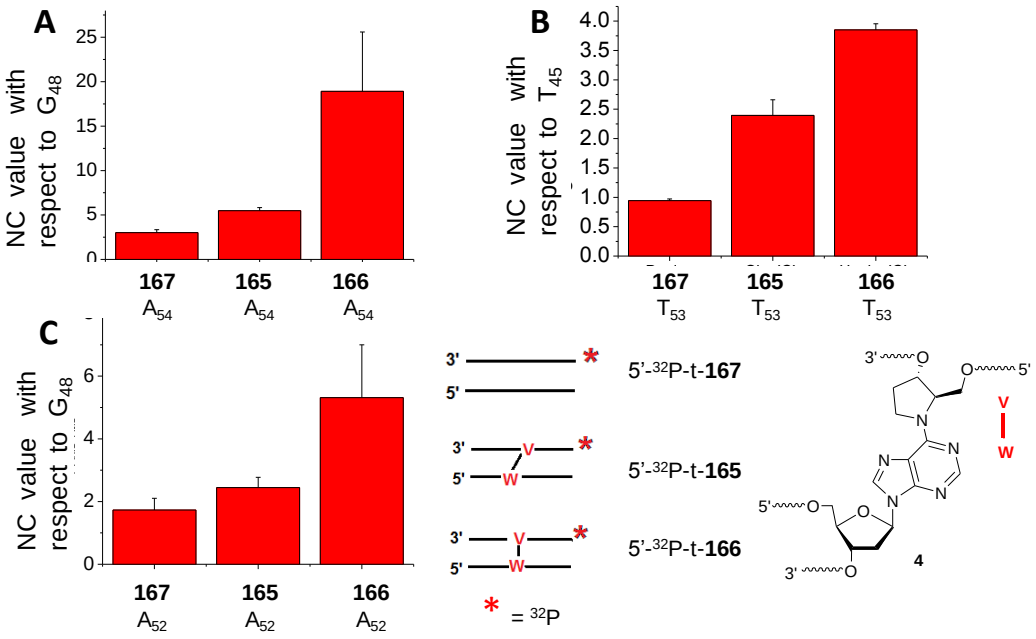


Figure 65. Comparison of cleavage of nucleotides (as a function of distance from ICL) in 5'-^{32}P-t-**167** (control duplex), 5'-^{32}P-t-**165** (observed ICL), and 5'-^{32}P-t-**166** (unobserved ICL) for DEPC and $KMnO_4$ reaction. (A) NC value for DEPC reaction of A_{54} for 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166** with respect to G_{48}. (B) NC value for $KMnO_4$ reaction of T_{53} from 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166** with respect to T_{45}. (C) NC value for DEPC reaction of A_{52} from 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166** with respect to G_{48}.

as a quantitative measure of their distortion. The NC value of A_{54} for 5'-^{32}P-t-**166** (unobserved ICL, NC = 18.91 ± 6.67) was the highest among the three (Figure 65A and Table 10). The ICL-induced distortion was prevalent in the next nucleotide, T_{53} (one nucleotide away from ICL) of 5'-^{32}P-t-**165** and 5'-^{32}P-t-**166**, which was sensitive towards $KMnO_4$ (App Figure 65). Comparison of the peak intensities in the histograms and the NC values (reference nucleotide T_{45}) indicated greater sensitivity for T_{53} in unobserved ICL (5'-^{32}P-t-**166**, NC = 3.85 ± 0.87) than the observed ICL (5'-^{32}P-t-**165**, NC = 2.39 ± 0.27) (App. Figure 65, Figure 65B, and Table 10).

Higher peak intensity of the next nucleotide A_{52} (two nucleotides away from ICL) in the histogram compared to the background indicated ICL-influenced backbone distortion in 5'-^{32}P-t-**165** and 5'-^{32}P-t-**166** (App. Figure 63B and App. Figure 64A-D). The

NC value [$\frac{\text{(Band intensity of target nucleotide)}}{\text{(Band intensity of reference nucleotide)}}$, reference nucleotide G_{48}] of A_{52} indicated highest reactivity for the unobserved ICL (5'-^{32}P-t-**166**, NC = 6.67 ± 1.68)

Table 10. Chemical probing experiments on the C4-AP ICL analogue 4 for 5'-side of abasic site analogue containing strand ("top" strand)

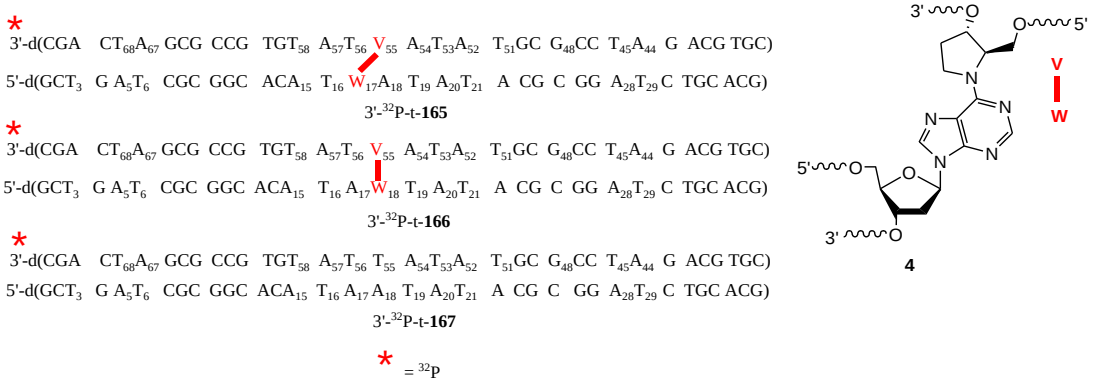
Nucleotide of interest in observed ICL (**165**), unobserved ICL (**166**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)
A_{54} of **167**	3.01± 0.34	G_{48} (DEPC)
A_{54} of **165**	5.47 ± 0.35	
A_{54} of **166**	18.91 ± 6.67	
T_{53} of **167**	0.94 ± 0.03	T_{45} ($KMnO_4$)
T_{53} of **165**	2.39 ± 0.27	
T_{53} of **166**	3.85 ± 0.17	
A_{52} of **167**	1.73 ± 0.37	G_{48} (DEPC)
A_{52} of **165**	2.44 ± 0.33	
A_{52} of **166**	6.67 ± 1.68	

(Figure 65C and Table 10). Treatment of 5'-^{32}P-t-**167** (control duplex), 5'-^{32}P-t-**165** (observed ICL), and 5'-^{32}P-t-**166** (unobserved ICL) with $KMnO_4$ did not demonstrate any hyperreactivity for T_{51} in the denaturing PAGE or histograms, indicating disappearing influence of ICL induced distortion (App. Figure 65A-D). The nucleotides 5' to T_{51} were not probed any further.

3.3.1.1.4. Reactivity of nucleotides on the 3'-side of the cross-linked abasic site containing strand

In the region 3' to the cross-linked abasic site (V_{55}) containing strand ("t" or "top"), the reactivity of the nucleotides up to T_{58} were investigated using their normalized cleavage or NC value [$\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]. Unlike the cross-linked abasic site in DOB ICL analogue **3**, its counterpart in **4** (V_{55}) was not

hyperreactive upon hydroxyl radical treatment of 3'-^{32}P-t-**165** (observed ICL) and 3'-^{32}P-t-**166** (unobserved ICL) (See Experimentals for their preparation.) (App. Figure 66A).



The nucleotide reactivity was uniform throughout the strand for control duplex (3'-^{32}P-b-**167**), observed ICL (3'-^{32}P-t-**165**), and unobserved ICL (3'-^{32}P-t-**166**) (App Figure 66B – D). We considered three different reasons to explain the absence of hyperreactivity at the abasic site. The possibility of multiple hit kinetics was ruled out due to the presence of ~70 – 75% starting material for 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166**. Secondly, the inherent lack of reactivity of abasic site analogue **4** seemed unlikely due to its resemblance to the abasic site in **3**. Therefore, we considered the third reason, absence of any solvent exposed sugar (the hydrogen of which is abstracted by hydroxyl radical) or its analogue in the backbone to be a plausible reason behind the lack of detectable hyperreactivity in the hydroxyl radical cleavage of 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166**. As a result, the reactivity of the cross-linked V_{55} of observed ICL (3'-^{32}P-t-**165**) and unobserved ICL (3'-^{32}P-t-**166**) could not be differentiated based on the NC value. Since the next nucleotide, T_{56}, was opposite a cross-linked A for 3'-^{32}P-t-**165** (observed ICL), we expected high sensitivity towards $KMnO_4$. Unlike its counterpart in the DOB "observed" ICL analogue (3'-^{32}P-t-**155**), T_{56} did not respond to $KMnO_4$ treatment of C4-AP observed ICL (3'-^{32}P-t-**165**) (App. Figure 67A and C). Instead, T_{56} in 3'-^{32}P-t-**165**

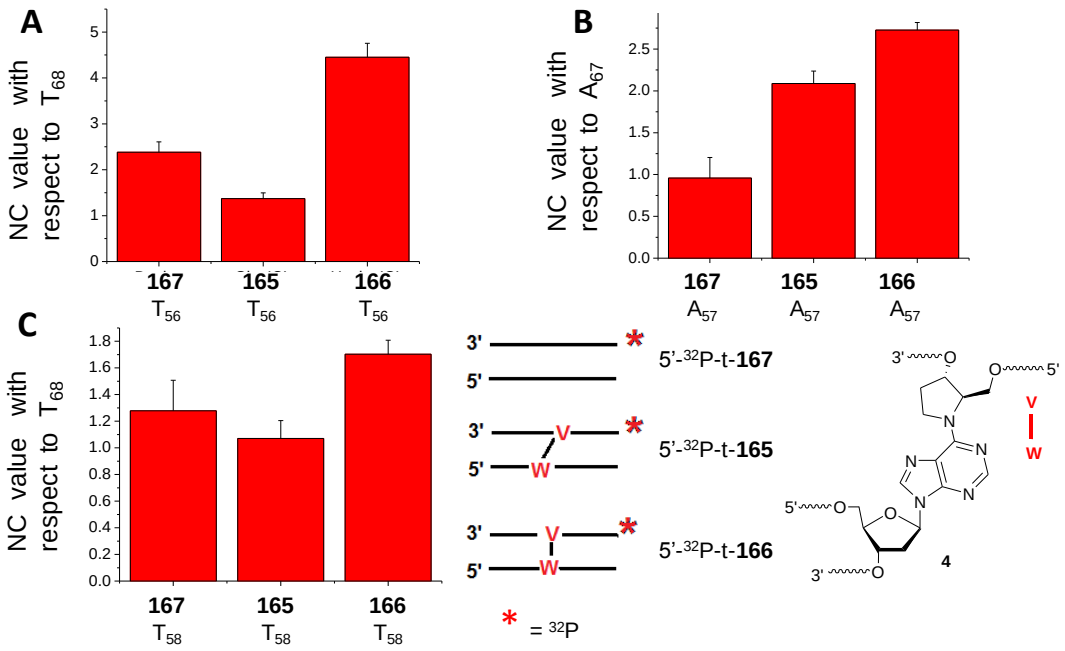


Figure 66. Comparison of cleavage of nucleotides (as a function distance from ICL) of in 3'-^{32}P-b-**167** (control duplex), 3'-^{32}P-t-**165** (observed ICL), and 3'-^{32}P-t-**166** (unobserved ICL) for DEPC and $KMnO_4$ reaction. (A) NC value for $KMnO_4$ reaction of T_{68} for 3'-^{32}P-b-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** with respect to T_{68}. (B) NC value for DEPC reaction of A_{57} from 3'-^{32}P-b-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** with respect to T_{67}. (C) NC value for $KMnO_4$ reaction of T_{58} from 3'-^{32}P-b-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** with respect to T_{68}.

(observed ICL) was more protected than the comparable nucleotide in control duplex (3'-^{32}P-t-**167**) (App. Figure 67A and B). In contrast, T_{56} in unobserved ICL (3'-^{32}P-t-**166**) was sensitive towards $KMnO_4$ (App. Figure 67A and D). Needless to mention, the NC value of T_{56} (4.45 ± 0.30, reference nucleotide T_{68}) was highest for unobserved ICL (3'-^{32}P-t-**166**) among the three substrates (Figure 66A and Table 11). The lack of reactivity at ICL-adjacent T_{56} in 3'-^{32}P-t-**165** (observed ICL) was more baffling due to the presence of hyperreactivity at A_{57} (one nucleotide away from ICL) in 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166** (App. Figure 68A – D). The reactivity of A_{57} was higher in unobserved ICL (3'-^{32}P-t-**166**, NC = 2.72 ± 0.09, reference nucleotide A_{67}) than in observed ICL (3'-^{32}P-t-**165**, NC = 2.08 ± 0.15) (Figure 66B and Table 11). The reactivity of the next nucleotide, T_{58} (two nucleotide away from ICL), followed the same trend as its neighbor, T_{56}, demonstrating

greater reactivity for the 3'-^{32}P-t-**166** (unobserved ICL, NC = 1.70 ± 0.10, reference nucleotide T_{68}) than the the 3'-^{32}P-t-**165** (observed ICL, NC = 1.07 ± 0.13) (App. Figure 67 and Figure 66C, and Table 11). The $KMnO_4$ treatment of 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166** did not yield any hyperreactivity for nucleotides 3'- to T_{58}, indicating disappeared influence of ICL induced distortion (App. Figure 67C and D). Nucleotides 3'- to T_{58} was therefore not investigated.

Table 11. Chemical probing experiments on C4-AP ICL analogue 4 for 3'-side of abasic site analogue containing strand ("top" strand)

Nucleotide of interest in observed ICL (**165**), unobserved ICL (**166**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}$]	Reference nucleotide (reagent)
A_{55} of **167**	1.00 ± 0.21	A_{67} (hydroxyl radical cleavage)
V_{55} of **165**	0.84 ± 0.38	
V_{55} of **166**	0.77 ± 0.10	
T_{56} of **167**	2.31 ± 0.22	T_{68} ($KMnO_4$)
T_{56} of **165**	1.36 ± 0.12	
T_{56} of **166**	4.45 ± 0.30	
A_{57} of **167**	0.95 ± 0.24	A_{67} (DEPC)
A_{57} of **165**	2.08 ± 0.15	
A_{57} of **166**	2.72 ± 0.09	
T_{58} of **167**	1.28 ± 0.23	T_{68} ($KMnO_4$)
T_{58} of **165**	1.07 ± 0.13	
T_{58} of **166**	1.70 ± 0.10	

3.3.1.1.5. Summary

To summarize, reaction with DEPC, $KMnO_4$ and hydroxyl radical revealed solvent exposed nucleotides on both strands of **165** and **166** (Figure 67). The distortion characterized by solvent exposed nucleotides was confined to within 3 – 4 nucleotides on each side from the ICL. Like DOB ICL analogue **3**, the reactivity of the nucleotides varied. At the same time, the distorted (solvent exposed) nucleotides in the top and

bottom strands of ICLs containing **4** were more evenly distributed than the DOB ICL analogue **3**. The varying degree of solvent exposure on the top and bottom oligonucleotide strands of containing **4** resembled the uneven nature of distortion detected in strands connected by asymmetric cross-linking agents like mitomycin C and acrolein.[88,92]

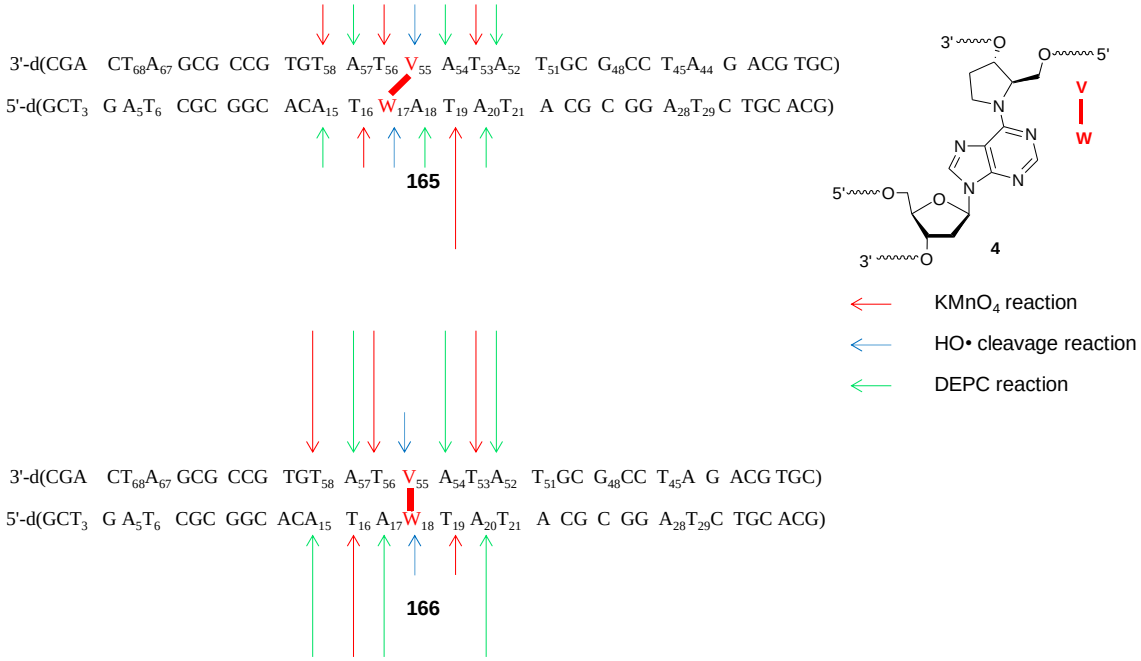


Figure 67. Histogram describing comparative analysis of chemical probing on oligonucleotides **165** and **166** containing DOB ICL analogue **3**. The length of the arrows represents the reactivity difference for a particular nucleotide (towards a specific chemical agent). The differences in length for two different nucleotides are not comparable.

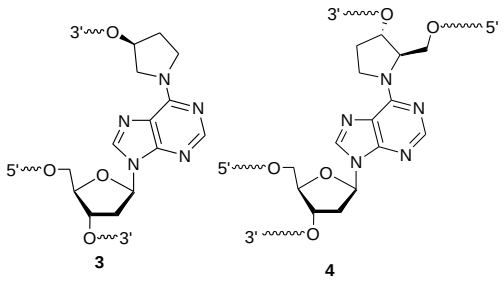
In general, greater chemical agent concentrations were required for inducing cleavage in observed (**165**) and unobserved ICL (**166**) compared to ICLs containing **3**. In the cross-linked A containing strand ("bottom" strand), sensitivity of nucleotides A_{15}, T_{16}, and A_{20} towards the respective chemical agents were higher for **166** (observed ICL) than **165** (unobserved ICL). In contrast, T_{19} demonstrated an unexplainable ~7 fold greater reactivity in observed ICL (**165**) than unobserved ICL (**166**). Hydroxyl radical cleavage did not detect any hyperreactivity in the cross-linked A in **165** or **166** (W_{17} and W_{18},

respectively). On the cross-linked abasic site analogue-containing strand ("top" strand), a greater degree of hyperreactivity was detected for nucleotides A_{52} to T_{58} in **166** (unobserved ICL) compared to **165** (observed ICL). Despite being present opposite a cross-linked A with possibly no H-bonding capability, T_{58} in observed ICL (**165**) was not sensitive towards $KMnO_4$. The lack of reactivity at T_{58} contradicted the sensitivity of the T_{47} opposite cross-linked A (X_{14}). Among the nucleotides of the top strand of **165** and **166**, no hyperreactivity was detected for the cross-linked abasic site (V_{55}) from the hydroxyl radical cleavage. This was in contrast to the hyperreactivity detected for cross-linked Y_{48} in **155** and **156** containing DOB ICL analogue **3**. Overall, the sensitivity of nucleotides in unobserved ICL (**166**) was generally higher than their counterparts in observed ICL (**165**), indicating the presence of a greater degree of distortion.

3.3.2.2. T_m analysis of substrates containing C4-AP ICL analogue (4)

3.3.2.2.1. Optimization of substrate sequence and salt condition for melting temperature analysis

The chemical probing experiments detected greater distortion in C4-AP ICL analogue **4** formed with opposing A (the "unobserved" type) than that with A opposite 3'-



adjacent T ("observed"). In order to validate our findings, the thermal stability of oligonucleotides containing the "observed" and "unobserved" type of **4** was investigated. Overall, the chemical probing experiments indicated less distortion in substrates

containing **4** than **3**. This prompted us to anticipate smaller differences in the T_m's of the "observed" and "unobserved" C4-AP ICL analogue **4** compared to those containing **3**. The sequence composition in ICL **4**, salt, and buffer conditions were optimized to maximize the difference between "observed" and "unobserved" ICLs. While we were certain about employing PIPES or phosphate buffer along with a monovalent cation for T_m experiments, the use of Mg^{2+} in the thermal melting of oligonucleotides was trickier.

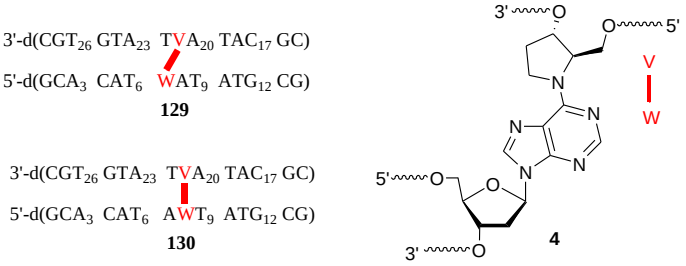
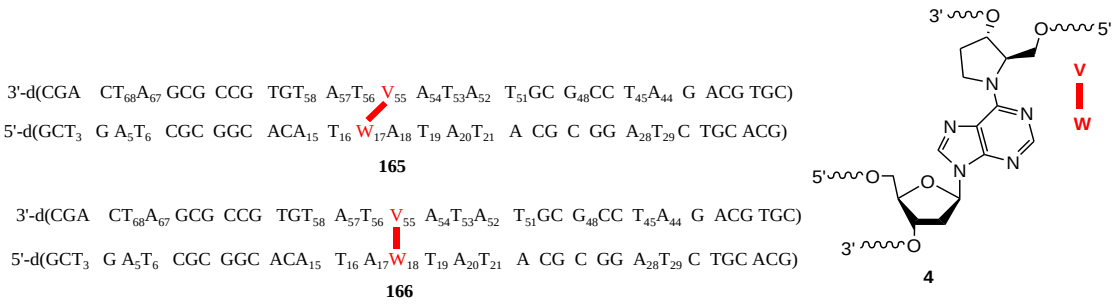


Figure 68. Oligonucleotides containing C4-AP ICL analogue **4** used for melting temperature analysis.

The correlation between thermal stability of duplexes and Mg^{2+} (0 – 600 mM) is not continuous.[161] The T_m initially rises with increasing Mg^{2+} concentration, reaching the highest thermal stability around 25 – 350 mM Mg^{2+}, then dropping steadily to 600 mM.[161] At the same time, the difference in T_m between substrates decreases from 0 to 350 mM, and then increases again up to 600 mM.[161] In addition, the difference in T_m behavior is dependent on the G-C content of the substrates as well.[161] Due to the nonlinear nature of the correlation, Tm experiments on **4** were initiated using Mg^{2+} concentration of 10 mM (this concentration was used for analyzing the thermal stability of **3)**.

In order to compare the thermal stability of the ICL formed with an opposing A (the "unobserved" type, **130**) and that with A opposite 3'-adjacent T ("observed", **129**), ICLs **129** and **130** were prepared using solid phase synthesis techniques developed earlier

(Figure 68). The core sequence flanking the ICL (**4**) in **129** and **130** (C_4 – T_{11}) was derived from C_{14} – T_{21} region encompassing the ICL in the chemical probing substrates



165 (observed ICL) and **166** (unobserved ICL), respectively. The remaining sequence composition and overall length of **129** and **130** were determined using theoretical T_m calculations using hypothetical hairpins **171** and **172** (Figure 69). Like DOB ICL analogue **3**, we anticipated the base paired regions on both sides of the ICL **4** to function like hairpins in T_m studies, elevating the T_m. The cross-linked region in **3** resembled a single hairpin, making prediction of its melting behavior simpler. In contrast, we envisioned ICL analogue **4** to consist of two separate hairpin stem-loop like structures. In principle, the A_6 – A_9 loop of the hairpins **171** and **172** (Figure 69) would represent the ICL-adjacent highly distorted solvent exposed region of **4** (e.g. $T_{16}W_{17}V_{55}T_{56}A_{57}$ and $T_{19}A_{18}W_{17}V_{55}A_{54}$ of **165**) detected in chemical probing. We anticipated that the vicinity of the A_5 – T_{10} pair in **171** with the loop would mimic the relatively low-level distortion detected in the A_{58} – T_{58} pair in **165** and **166**. Similar kind of correlation was expected of the T_5 – A_{10} in **172** and A_{20} – T_{53} in the cross-linked substrates (**165** and **166**). Assessment of the length of the stems of **171** and **172** (and corresponding length of arms on each side of **4** of the T_m substrates) was trickier. To the best of our knowledge, earlier studies on thermal melting of ICLs without a strand break (e.g. those formed from mitomycin C or psoralen)[38,79] did not address whether the melting of the two hairpin-like

structures therein comprised of one or two separate transitions. In the absence of such theory or experiment, we assumed the thermal melting of the two hairpin like structures in ICL **4** would behave as separate transitions. However, we assumed that similar

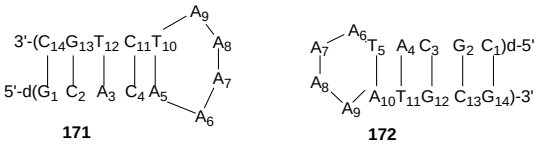
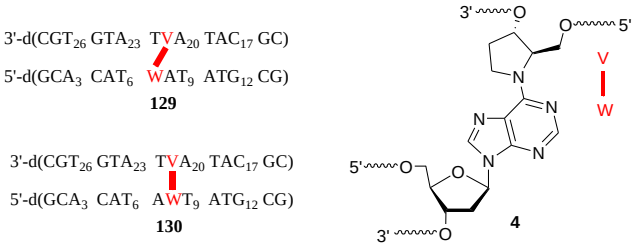


Figure 69. Hypothetical hairpin oligonucleotides used to estimate the T_m of cross-link substrates containing **4**.

G-C content on either side of **4** would bring their T_m's close enough so that both would be denatured within a similar temperature range. While the region $G_{11} – A_{15}$ of **165** was adapted as the stem of the hairpin **171** ($G_1 – A_5$), $A_{20} – G_{24}$ of **165** was slightly modified as the $A_{10} – G_{14}$ of **172** to achieve a similar G – C content. The stems of **171** and **172** ($G_1 – A_5$ and $A_{10} – G_{14}$, respectively) would resemble the base paired region distant from ICL **4** (e.g. $G_{11} – A_{15}$ and $A_{20} – G_{24}$ of **165** and **166**) which lacked any influence of distortion and sensitivity towards chemical agent. Next, we carried out T_m calculations on hypothetical hairpins **171** and **172** (Figure 69). At 100 mM Na^+ (pH 7.0), 10 mM Mg^{2+} and oligonucleotide concentration 1 μM, the melting temperatures of **171** and **172** were estimated to be ~58 °C and ~62 °C using an online T_m calculator (https://www.idtdna.com/calc/analyzer). If our analogy of ICL **4** structure with two hairpins is correct, the proximity of estimated T_m values would cause the melting of both regions of **4** to occur within a common temperature range. From our experience with DOB ICL analogue (**3**), we anticipated that thermodynamic stability of **129** (observed ICL) and **130** (unobserved ICL) can be distinguished based on their T_m in this temperature range. If the hypothetical and experimental T_m would be found close enough, this method can be applied to estimate the T_m of any cross-link with available structural

data and formed without an adjacent strand break, such as those formed from mitomycin C or psoralen.[38,79]

3.3.2.2.2. Calculation of extinction coefficients of oligonucleotides containing C4-AP ICL analogue 4



The extinction coefficient (Ɛ) of oligonucleotides **129** and **130** at 260 nm were calculated based on the nearest-neighbor model and a list of corrected extinction coefficients of nucleotides (http://www.owczarzy.net/extinctionDNA.htm).[156–158] The list and methodologies from nearest-neighbor model are combined in a standard extinction coefficient calculator (http://biophysics.idtdna.com/UVSpectrum.html). To simplify the calculation, we assumed the hairpin like regions of **129** and **130** would remain base paired even in the absence of salt. Based on this premise, we assumed nucleotides G_1 – W_7 in **129** to be a duplex, with W being A. The cross-linked abasic side analogue on top strand was not expected to contribute in the extinction coefficient. Similarly, nucleotides G_1 – A_7 in **130** were assumed to form a duplex as well. Both duplexes are estimated to have Ɛ = 112770.4 $L \bullet mol^{-1} \bullet cm^{-1}$ (based on the online Ɛ calculator). Due to their location opposite the cross-linked abasic site analogue, A_8 in **129** and W_8 in **130** were considered

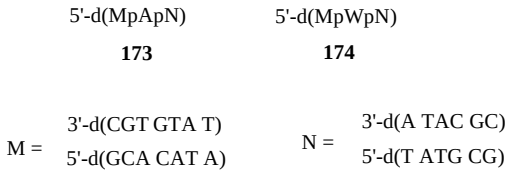


Figure 70. Hypothetical oligonucleotides used for estimation of extinction coefficients of **129** and **130**.

single stranded residues without any base pairing or stacking stabilization. On the other hand, nucleotides $T_9 – G_{14}$ were assumed to form a duplex in both the substrates with an estimated extinction coefficient 97420.6 $L•mol^{-1}•cm^{-1}$ (based on the online Ɛ calculator). Next, we hypothesized **129** and **130** as two linear oligonucleotides 5'-d(M_pA_pN)-3' (**172**) and 5'-d(M_pW_pN)-3' (**173**). M and N, both imaginary nucleotides, represented the duplex regions $G_1 – W(A)_7$ and $T_9 – G_{14}$ of **129** and **130**, respectively (Figure 70). The A and W in **172** and **173** were representative of unpaired A_8 and W_8 in **129** and **130**, respectively. According to the nearest-neighbor model, the extinction coefficient (Ɛ) of **172** and **173** (and therefore of **129** and **130**) were calculated as follows,

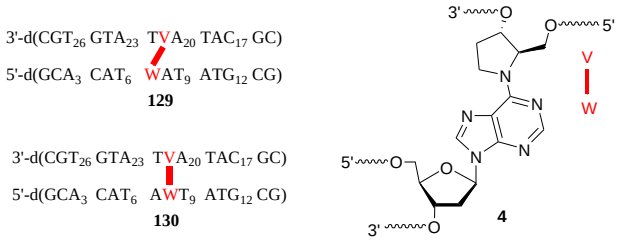
$$Ɛ_{5'\text{-d(MpApN)-3'}} = Ɛ_{dMpdA} + Ɛ_{dApdN} – Ɛ_{pdA}$$

$$= (Ɛ_{dM} + Ɛ_{pdA}) + (Ɛ_{dA} + Ɛ_{pdN}) – Ɛ_{pdA}$$

$$= Ɛ_{dM} + Ɛ_{pdA} + Ɛ_{dN}$$

Based on the above equation, the extinction coefficient of **129** and **130** was estimated to be 225591 $L•mol^{-1}•cm^{-1}$. We anticipated ~5% error in this value due to the direct addition between $Ɛ_{dM}$ (or $Ɛ_{dN}$) and $Ɛ_{dA}$ during calculation of $Ɛ_{dMpdA}$ and $Ɛ_{dApdN}$.

3.3.2.2.3. Melting temperature study of oligonucleotides containing C4-AP ICL analogue 4



We carried out a UV-melting study on **129** (observed ICL) and **130** (unobserved ICL) [DNA concentration 1 µM, PIPES (pH 7.0) 10 mM, Mg^{2+} 10 mM, Na^{+} 100 mM].

Both substrates melted around 60 °C (Figure 71 and Table 12). The melting temperatures were in accordance with the estimated T_m of hypothetical oligonucleotides (**170** and **171**) used to design the sequence composition of **129** and **130**. For **129** and **130**, the T_m values for the forward experiments (temperature gradually increased from 20 °C to 85 °C,

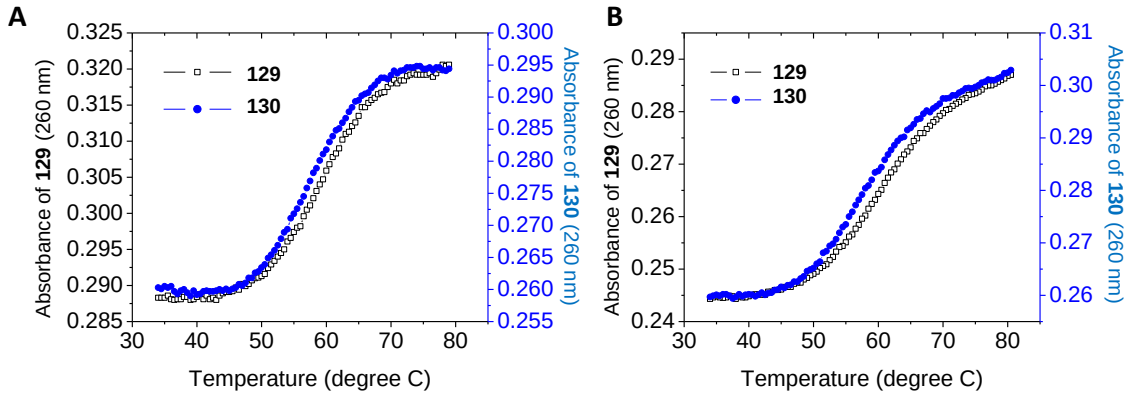


Figure 71. Thermal denaturation analysis of oligonucleotides containing C4-AP ICL analogue **4**. (A) Absorbance (260 nm) vs temperature profile of cross-linked substrate **129** (observed ICL) and **130** (unobserved ICL) during forward experiment (temperature increased from 20 to 85 °C. DNA concentration 1 μM, PIPES, pH 7.0, 10 mM, Mg^{2+} 10 mM, Na^{+} 100 mM. (B) Absorbance (260 nm) vs temperature profile of cross-linked substrate **129** and **130** during reverse experiment (temperature decreased from 85 to 20 °C. DNA concentration 1 μM, PIPES (pH 7.0) 10 mM, Mg^{2+} 10 mM, Na^{+} 100 mM.

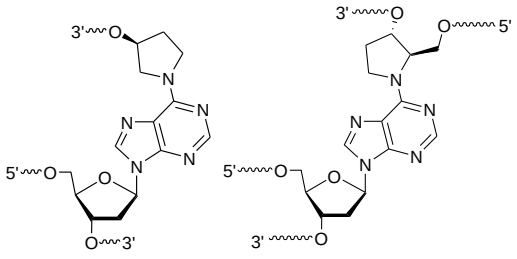
Figure 71A) were slightly lower than the respective reverse experiment (temperature gradually decreased from 85°C to 20 °C, Figure 71B and Table 12). We considered this difference to be an artifact since the values were within error limits of each other. An important difference detected in the T_m profiles of **3** and **4** was the broadness of the melting transitions. Compared to ~18 – 20 °C melting transition range observed for DOB ICL analogues (**3**), the change in UV absorption for **4** occurred over a broader region (~23 – 27 °C). We rationalized the increase in melting temperature range by the slightly different G-C distribution in the two duplex arms (G_1 – T_6 and T_9 – G_{14}) of **129** and **130**. As a result, the two hairpin-like regions with the different G-C content would have slightly different T_m, leading to overlapping and broader melting transitions. Broad UV

melting transitions are also reported during thermal denaturation of mitomycin C[41] and N4C-ethyl-N4C ICLs.[91] Most importantly, we detected consistently greater T_m values for observed ICL (**129**) over unobserved ICL (**130**) for both forward and reverse experiments (Table 12). This validated the greater thermodynamic stability for "observed" cross-linking with A opposite a 3'-adjacent T (as in **129**) over "unobserved" ICL formationwith an opposing A (as in **130**). The observations also corroborated with the chemical probing results, in which oligonucleotides containing ICL **4** with an opposing A were more distorted (characterized by solvent exposure). In addition, the difference in T_m between "observed" ICL **129** and "unobserved" **130** was much less compared to that between the observed and unobserved DOB ICL analogue **3** (**160** and **161**, $\Delta T_m \sim 4$ °C for the hairpin like cross-linked region, Table 7). This was despite the fact that the length of the duplex region was shorter in the C4-AP ICL analogue **4** than for **3**. The smaller difference in C4-AP ICL T_m's is consistent with the lower distortion detected for oligonucleotides containing **4** during chemical probing.

Table 12. Melting temperature of oligonucleotides containing C4-AP ICL analogue 4 (averaging is based on at least three replicates)

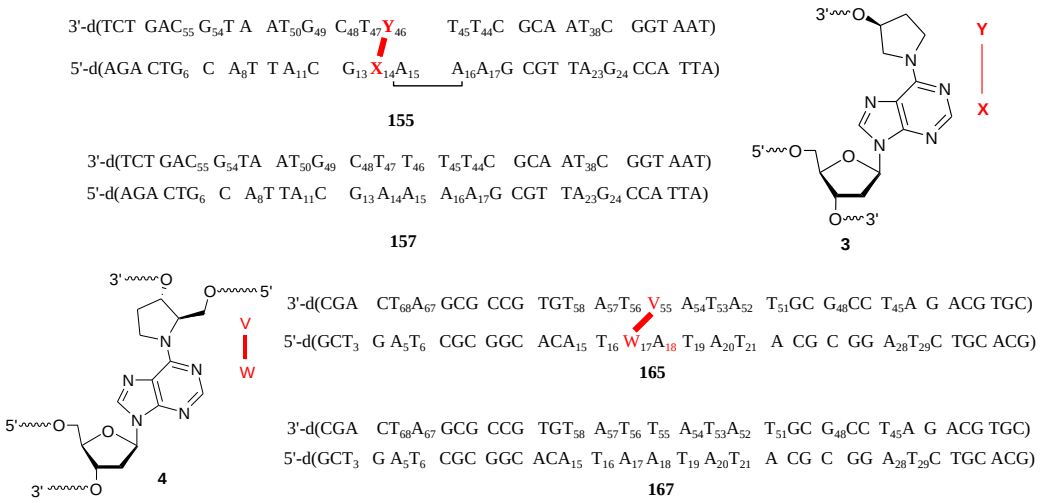
Substrate	T_m in forward experiment (°C)	T_m in reverse Experiment (°C)
129	59.07 ± 0.23	60.09 ± 0.66
130	57.42 ± 0.41	57.91 ± 0.32

3.3.1.1.5. Correlation between nucleotide solvent exposure, strand break and UvrABC incision of 3 and 4



3 4

Earlier, we hypothesized that the lower sensitivity of oligonucleotides containing C4-AP ICL analogue (**4**) towards chemical probing agents compared to those containing the DOB ICL analogue (**3**) substrates were due to less distortion in the former. In addition, ICL **4** had smaller differences in T_m's between the "observed" (adduct formation with A opposite 3'-adjacent T) and "unobserved" (ICL with opposing A) cross-links than the analogous substrates containing **3**.



While both observations support the notion of less distortion in **4**, we attempted to validate it with a direct comparison between **3** and **4**. Previously, nucleotide reactivity towards chemical agents were compared between a mitomycin C ICL and the corresponding un-cross-linked duplex.[40] The comparison detected the distribution of

solvent exposed nucleotides and correlated well with the NMR derived structure of mitomycin C ICL.[88] Intrigued, I compared the solvent exposure of nucleotides (characterized by sensitivity towards chemical agents) in cross-linked substrates **155** and **165** with a reference consisting of the respective un-cross-linked duplexes (**157** and **167**). Oligonucleotides **155** and **165** were chosen because their cross-link connectivity (the "observed" type) matched those used earlier in UvrABC nucleotide excision repair experiments. The ratio of the NC value of a particular nucleotide from ICL to that of the same nucleotide in the corresponding un-cross-linked duplex was estimated as a measure of solvent exposure. Only those nucleotides that showed signs of hyperreactivity in the chemical probing experiments were chosen for this analysis. The ratios of NC values between the duplex and ICL were plotted against the same scale for nucleotides in top (having the abasic site analogue) and bottom strand (containing the cross-liked A) (Figure 72). The comparison of the ratio of NC values between **155** and **165** clearly indicated a high degree of distortion around the ICL **3** in **155**, especially in the top strand (Figure 72A). Among the nucleotides demonstrating very high degree of distortion were cross-linked abasic analogue Y_{46}, T_{47}, T_{45} and T_{44}. In addition, A_{15}, A_{16} and A_{17} were among hyperreactive sites on the other strand. On the other hand, the sensitivity of distorted nucleotides of **165** (containing C4-AP ICL analogue **4**) towards chemical agents were more evenly distributed with the exception of T_{19}, which was detected with a surprisingly high reactivity. With the exception of T_{19}, none matched the extreme solvent exposure demonstrated by nucleotides encompassing DOB ICL analogue **3** in **155**.

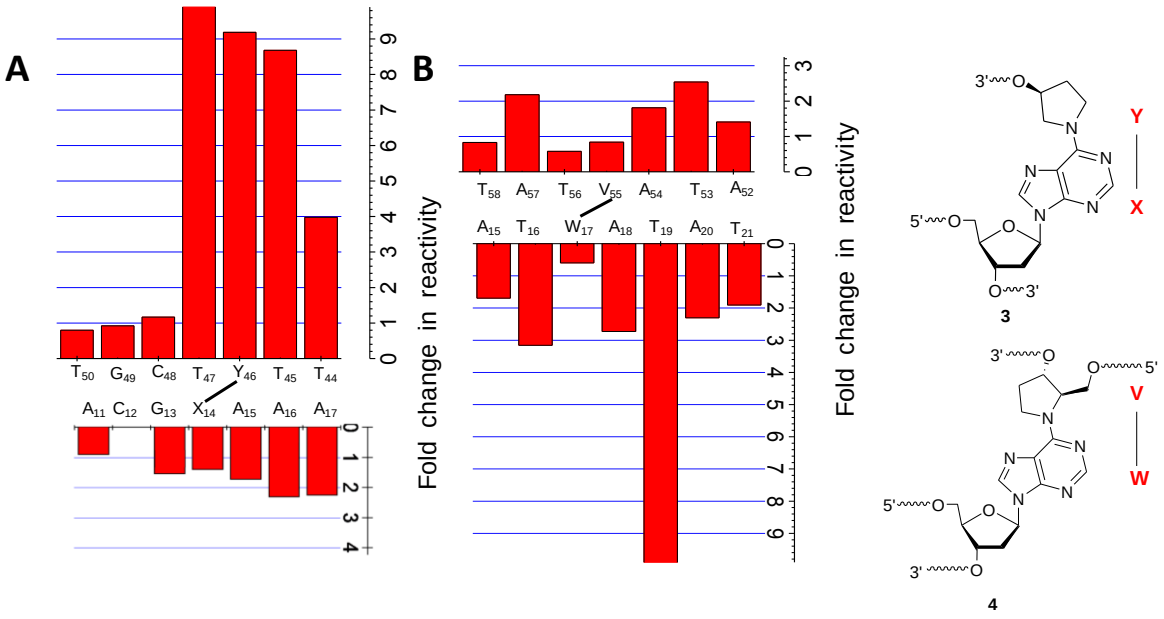


Figure 72. The reactivity of nucleotides in cross-linked substrates **155** (observed ICL containing DOB ICL analogue **3**) and **165** (observed ICL containing C4-AP ICL analogue **4**). (A) Comparative reactivity of nucleotides in **155** with reference to **157**. The reactivity of C_{12} was not included due to lack of hydroxylamine probing on the C_{12} of **157**. (B) Comparative reactivity of nucleotides in **165** with reference to **167**.

One possible reason behind the greater distortion in substrates containing DOB ICL analogue **3** could be the presence of the strand break adjacent to the ICL, which provides additional flexibility. The presence of a continuous helical backbone in **4** may help contain distortion. Despite the ICL induced distortion, the continuous backbone holds the adjacent

3'-d($T_{45}T_{44}$C GCA AT_{38}C GGT AAT)
5'-d(AGA CTG_6 C A_8T TA_{11}C $G_{13}A_{14}A_{15}$ $A_{16}A_{17}$G CGT $TA_{23}G_{24}$ CCA TTA)
158

nucleotides together, preventing disruption of stabilizing forces like base stacking and H-bonding. In contrast, a strand break itself introduces a number of solvent exposed nucleotides in the terminus adjacent to the nick (e.g. detection of hyperreactivity at T_{45} and T_{44} of **158**, App Figure 53). In combination with ICL induced distortion, an adjacent

strand break as in **3** could thus produce greater disruption in base stacking and greater solvent exposure in the region flanking the cross-link. According to the current model of UvrABC damage detection, the bulged backbone of a damaged DNA backbone causes steric clash with the protein wall of a narrow tunnel in $UvrA_2B_2$ complex. This steric clash distinguishes a damaged DNA from an undamaged one without any bulged backbone, which fits well within the tunnel.[110] Greater distortion in a substrate containing DOB ICL analogue **3** would result into greater steric clash in the protein tunnel of $UvrA_2B_2$, inducing more efficient preincision complex formation[124] and higher incision yield.

While the greater distortion in DOB ICL analogue **3** was consistent with a higher incision yield by UvrABC endonuclease, the distribution of solvent exposed nucleotides therein could not explain the regioselective nature of the observed incision. Earlier studies have correlated helical instability with greater incision yield,[102] possibly due to the ease of local unwinding for a thermodynamically less stable region. Although the hairpin-like cross-linked region in **3** had a number of solvent exposed nucleotides including the abasic site analogue, the high T_m values indicated significant thermodynamic stability. In contrast, low T_m of the un-cross-linked region revealed a considerably less thermodynamically stable structure. Since UvrABC incision occurred on the strand not carrying the nick on the un-cross-linked region of **3** (Figure 43),[123] the distorted ICL was presumably the alluring reason for UvrA binding and preincision complex formation. The combination of local distortion in the un-cross-linked region (visualized in the form of solvent exposed nucleotides) with the low structural stability is possibly causing regioselective UvrC incision.

3.3.3. Summary

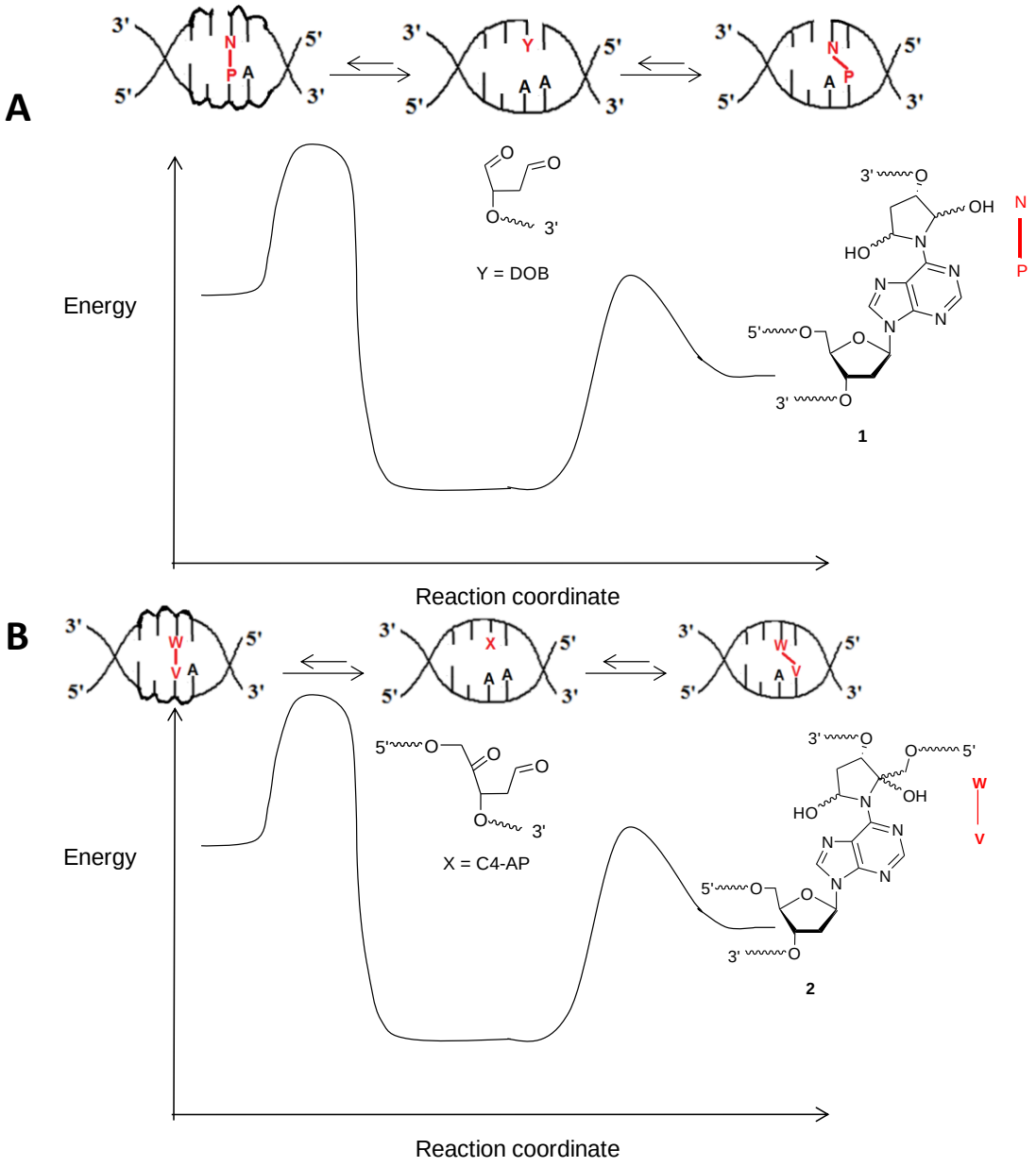


Figure 73. Energy dynamics of site selective DOB (A) and C4-AP ICL (B) formation.

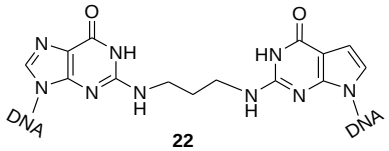
In conclusion, the "observed" (ICL with A opposite 3'-adjacent nucleotide) and "unobserved" (ICL with opposing A) type of cross-link formation in DOB (**3**) and C4-AP ICL (**4**) analogues were successfully probed using chemical agents. Quantification of solvent exposure of a nucleotide during chemical probing is traditionally estimated using

its percent cleavage intensity in gel compared to that of the unreacted starting material.[147,148] However, we identified that such computation would be susceptible to human errors such as variation in reaction time, and concentration of reagents. Therefore, we employed an internal normalization of cleavage band volume of the nucleotide of interest (closer to ICL) by taking its ratio with that of a potentially unperturbed one (far from ICL). We anticipated that such "normalized cleavage" or NC value $\left[\frac{(\text{Cleavage band volume of nucleotide of interest})}{(\text{Cleavage band volume of reference nucleotide})}\right]$ would be relatively insulated from the above-mentioned human errors, at least to some degree. Computation of NC values for **3** and **4** determined that the distortions were contained within 3 – 4 nucleotides from the ICL in both directions. However, an uneven distribution of deformity in the cross-linked strands was detected for DOB ICL analogue (**3**) and C4-AP ICL analogue (**4**). The range and distribution of ICL induced distortion (for both **3** and **4**) were similar to that for other asymmetric ICLs.[88,92]

Comparison of the NC values obtained from the "observed" and "unobserved" types of ICLs for **3** and **4** revealed that the nucleotides in the "unobserved" substrates were consistently associated with a greater degree of solvent exposure. This implied that the "unobserved" (cross-linking with opposing A) type of ICLs were more distorted than their "observed" (cross-linking with A opposite 3'-adjacent T) counterparts. This observation was corroborated by UV thermal melting studies, where the T_m of the "unobserved" type of cross-links were lower than their "observed" counterparts for **3** and **4**.

In contrast to NMR, chemical probing cannot be used to detect specific reasons behind solvent exposure, such as changes in nucleobase orientation (*syn* or *anti*), non-

native deoxyribose conformation, or change in base pair geometries. The advantage of chemical probing lies in the use of significantly smaller oligonucleotide quantities, and relatively straightforward qualitative detection of distorted nucleotides. The observation of greater distortion in the "unobserved" type of ICL is consistent with NMR and T_m studies on trimethylene ICL (**22**), the reduced version of acrolein ICL, where the "unobserved" ICL in 5'-(GC) sequence had greater distortion than the "observed" 5'-(CG) ICL.[93]



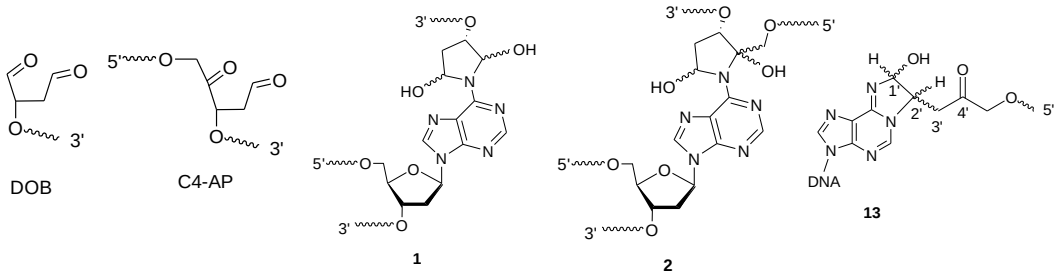
22

The observation of higher solvent exposure in nucleotides of "unobserved" **3** and **4** is consistent with the hypothesis of greater degree of distortion present during cross-linking with opposing A. Adduct formation with an opposing A thus could be associated with an energetically less stable intermediate (associated with an unfavorable transition state) than that with an A opposite a 3'-adjacent T (Figure 73A and B). Since the generation of **1** and **2** from their respective lesions are reversible processes, ICL formation with opposing A should have a lower K_{eq}, instead favoring the more stable "observed" adduct formation.

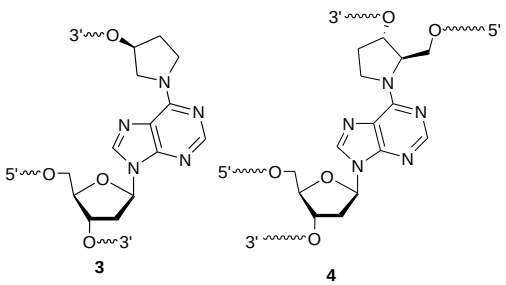
4. Conclusion

C4'-abasic site (C4-AP) and dioxobutane (DOB) lesions are generated by oxidative assaults on DNA. These lesions form interstrand cross-links (ICLs) with the N6 amine of A. The DOB, a lesion generated concomitantly with a 5'-strand break, forms a short-lived ICL (**1**, half life 10 h).[9] The C4-AP initially forms an unstable kinetic ICL (**2**,

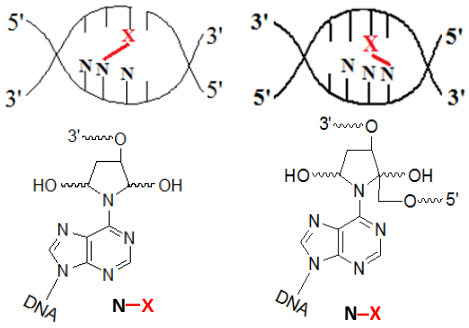
half life 3.1 h), which reverts to C4-AP, ultimately undergoes a strand scission and forms a stable thermodynamic ICL (**13**).[31] Although both ICLs (**1** and **2**) are unstable, they are



capable of disrupting essential cellular processes in their short lifetime. Comprehension of their biochemical effect thus remains necessary. In this report, we described synthetic strategies for making stabilized analogues (**3**, **4**) of DOB and C4-AP kinetic interstrand cross-links (**1** and **2**).[9,31] Orthogonal chemistry compatible with solid phase DNA synthesis was developed for synthesis of the analogues (**3** and **4**).[162] In combination with enzymatic ligation, we synthesized long oligonucleotides (60 bp) suitable for bacterial nucleotide excision repair studies. In principle, the analogues can be used to make even longer substrates suitable for mammalian repair studies[163] as well. The synthetic strategies have broad applicability and are not restricted by sequence constraints. They also can be used to synthesize other cross-links (both DNA and RNA), and branched oligonucleotides.



Long oligonucleotides containing the stabilized ICL analogues (**3** and **4**) were subjected to UvrABC, bacterial nucleotide excision repair protein responsible for cross-link removal.[141] The C4-AP cross-link analogue **4** was incised on both the cross-linked A and abasic site analogue containing strand. The cleavage yield was low (maximum 30% in 8 h), indicating C4-AP to be a persistent block in essential cellular processes, such as transcription and replication. In contrast to previous findings involving A-T and mitomycin C ICLs,[20,21] no double strand break (DSB) was detected from repetitive incisions on both strands. Treatment of the DOB ICL analogue (**3**) with UvrABC produced exclusive incision on the strand containing cross-linked A, producing a DSB. Previously, treatment of C4-AP thermodynamic ICL (also contained a strand break) with UvrABC has resulted formation of DSB with 15% of the total incision yield.[22] To the best of our knowledge, DOB ICL analogue **3** is the only substrate that exclusively yields a DSB upon UvrABC treatment.



DOB ICL (**1**) C4-AP kinetic ICL (**2**)

The DOB and C4-AP kinetic ICLs form in a site selective and reversible manner, with the N6 amine of an A opposite the 3'-adjacent nucleotide, while almost no ICL formation was observed with A opposite the lesions themselves. The measurement of distance between N6 amine nitrogen of the As and the C1' of aldehyde of DOB or C4-AP

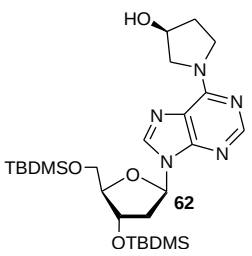
lesions using a Spartan® model did not justify the high selectivity. In addition to distance, we hypothesized that the site selectivity could be driven by the relative stabilities of the two type of ICLs, with the favored one being less distorted. Based upon this hypothesis, any potential ICL with an opposing A ("unobserved") should be more distorted and will have a more solvent exposed helical backbone than the ICL formed with an A opposite the 3'-adjacent T ("observed"). Utilizing the strategies developed for making C4-AP and DOB ICL analogues (**3** and **4**), cross-links with both connectivities ("observed" and "unobserved") were synthesized. All four sets of cross-links (the "observed" **3** and **4**, and the "unobserved" **3** and **4**) were subjected to nucleotide specific chemical reagents that predominantly react with solvent exposed nucleobases.[146–151] For both **3** and **4**, the nucleotides in cross-links with "unobserved" connectivity were more reactive than those with "observed" ICLs, indicating greater solvent exposure in the former. This observation was further corroborated by melting temperature analysis of four sets of cross-links, where **3** and **4** with "unobserved" type cross-links had lower T_m's.

5. Experimentals

General Methods. Solvents used in reactions were purified and dried (using CaH_2 or Na/benzophenone) by distillation before use. Reagents were purchased from commercial sources and were used without further purification. Reactions were carried out under a positive pressure of argon atmosphere and monitored by TLC on silica gel G-25 UV254 (0.25 mm). Spots were detected using UV light and/or by charring with a solution of either ammonium molybdate, ceric ammonium sulfate in water and H_2SO_4, or p-anisaldehyde in ethanol and H_2SO_4. Flash chromatography was performed on silica gel 60 (40-60 μm). The ratio between silica gel and crude product ranged from 100:1 to 20:1 (w/w).

Oligonucleotides were synthesized via automated DNA synthesis on an Applied Biosystems model 394 instrument. Commercially available DNA synthesis reagents including phosphorylation reagent was purchased from Glen Research. Phosphoramidites **66a**, **66b**, **113a**, and **113b** were dried for 12 h in a lyophilizer prior to the experiment. Oligonucleotides containing cross-link analogues were deprotected using concentrated NH_4OH at 55 °C for 12 h. Oligonucleotides containing native nucleotides only were deprotected using 1:1 mixture of aqueous methylamine (40%)-concentrated concentrated NH_4OH at 55 °C for 1 h. Oligonucleotides were purified by 20% denaturing polyacrylamide gel electrophoresis (PAGE), isolated by the crush and soak method, and desalted using C-18-Sep-Pak cartridges (Waters). Oligonucleotides containing modified nucleotides were characterized by MALDI-TOF MS or ESI-MS. Oligonucleotides were 5'-^{32}P-labeled by T4 polynucleotide kinase (New England Biolabs) and γ-^{32}P-ATP (Perkin-Elmer). Radiolabeling of the 3'-termini was done using α-^{32}P-cordycepin

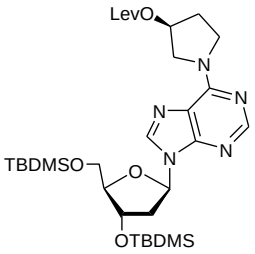
triphosphate (Perkin-Elmer) and terminal deoxynucleotidyl transferase (NEB). Radioisotopic "fill in" experiments at 3'-termini were carried out with Klenow (exo-) (NEB) and α-^{32}P-dCTP (Perkin-Elmer). Experiments involving radiolabeled oligonucleotides were analyzed following PAGE using a Storm 840 phosphorimager and Imagequant TL software. The band intensities of the target and reference nucleotides in chemical probing experiments was analyzed using the 1D gel analysis feature of ImageQuant TL, along with the "Rolling Ball" automated background correction. RSaI, Taq$^{\alpha}$I, CviQI, Fnu4HI, BsrBI, Taq$^{\alpha}$I, HpyCH4V, NEB 2.0, NEB 3.1, and CutSmart buffers were from NEB.



Synthesis of 62. The hydrochloride salt of **62** (667 mg, 5.5 mmol) was added to a stirred solution of **61** (1.6 g, 2.7 mmol) in THF (50 mL), followed by addition of Cs_2CO_3 (4.2 g, 13.5 mmol). The reaction was allowed to stir for 3.5 h at room temperature at which time the solvent was evaporated in vacuo. The crude material was dissolved in DCM (100 mL) and washed with saturated sodium bicarbonate (70 mL). The organic layer was collected and dried over sodium sulfate and concentrated. Flash chromatography (70% EtOAc in DCM to 100% EtOAc) yielded 11.2 g of **63** (82%). R_f (5% MeOH in DCM) = 0.3. ^{1}H NMR ($CDCl_3$): δ 8.3 (s, 1H), 8.0 (s, 1H), 6.41 (t, 1H, J = 6.4 Hz), 4.60 (m, 1H), 4.56 (m, 1H), 3.84-4.54 (m, 5H), 3.81 (dd, 1H, J = 3.9, 11 Hz), 3.72 (dd, 1H, J = 3.7, 11 Hz), 3.05 (bd s, 1H), 2.59 (dt, 1H, J = 5.6, 12.2 Hz), 2.38 (m, 1H), 2.13 (m, 2H), 1.89 (bd s, 1H), 0.88 (s, 18H), 0.074 (s, 6H), 0.070 (s, 6H). ^{13}C NMR ($CDCl_3$): δ 153.2, 152.7, 149.8, 137.2, 120.8, 87.8, 84.2, 72.3, 63.1, 45.8, 41.1, 26.1, 25.9, 18.6, 18.2, 8.7, -4.5, -4.6, -5.2, -5.3. IR (neat): 3300 (bd), 2951, 2856, 1591, 1471, 1252, 1105, 1031 cm^{-1}.

HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{26}H_{48}N_5O_4Si_2$ 550.3239, found 550.3250.

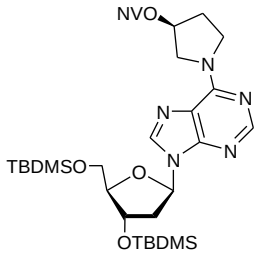
Synthesis of 63a. Levulinic acid (604 mg, 529 µL, 5.1 mmol), DMAP (731 mg, 5.9 mmol), and EDCI (989 mmol, 5.2 mmol) were added to a stirred solution of **62** (405 mg, 0.7 mmol) in DMF (9 mL). The reaction mixture was stirred for 4 h at room temperature under inert atmosphere. The resulting light orange mixture was diluted with EtOAc (80 mL) and then washed with water (50 ml), saturated sodium bicarbonate (50 mL), and brine solution (50 mL). The organic layer was dried over sodium sulfate, concentrated in vacuo and subjected to flash chromatography using 30% EtOAc in DCM to obtain the desired product (445 mg, 93%). R_f (30% EtOAc in DCM) = 0.31. ^{1}H NMR ($CDCl_3$): δ 8.32 (s, 1H), 7.97 (s, 1H), 6.42 (t, 1H, J = 6 Hz), 5.45 (m, 1H), 4.56 (s, 1H), 3.97-4.54 (m, 5H), 3.76 (dd, 1H, J = 4.6, 10.8 Hz), 3.72 (dd, 1H, J = 2.5, 10.8 Hz), 2.69 (m, 2H), 2.59 (m, 1H), 2.52 (m, 2H), 2.37 (m, 1H), 2.19 (m, 2H), 2.12 (s, 3H), 0.88 (s, 18H), 0.07 (s, 6H), 0.04 (s, 6H). ^{13}C NMR ($CDCl_3$): δ 206.4, 172.3, 153.1, 152.8, 150.1, 137.5, 120.8, 87.8, 84.1, 72.2, 62.9, 41.1, 37.9, 29.9, 28.2, 26.1, 25.8, 18.5, 18.1, -4.6, -4.7, -5.3, -5.4. IR (neat): 2852, 2926, 2856, 1720, 1591, 1471, 1156, 1093, 834, 777 cm^{-1}. HRMS (ESI/APCI) *m/z* calculated for $[M + H]^+$ $C_{31}H_{54}N_5O_6Si_2$ 648.3607, found 648.3616.



63a

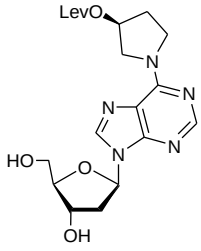
Synthesis of 63b. A mixture of **62** (304 mg, 0.55 mmol), *o*-nitroveratryloxycarbonyl chloride[164] (615 g, 2.22 mmol) and dimethylaminopyridine (549 mg, 4.43 mmol) were stirred in DCM (6 mL) at room temperature for 2 h, at which time the solution was diluted with 50 mL DCM. The diluted contents were washed with saturated sodium

bicarbonate (2 × 50 mL) followed by washing with brine (50 mL). The combined aqueous layers were washed with DCM (50 mL) and the combined organic layers were dried over sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in DCM. It was purified by flash chromatography (10% EtOAc in DCM to 60% EtOAc in DCM) to yield the desired product as a white solid (401 mg, 92%). R_f (50% EtOAc in DCM): 0.6. ^{1}H NMR ($CDCl_3$): δ 8.35 (s, 1H), 7.99 (s, 1H), 7.72 (s, 1H), 7.02 (s, 1H), 6.45 (t, 1H, *J* = 8 Hz), 5.59 (s, 2H), 5.45 (m, 1H), 4.60 (m, 1H), 4.01- 4.38 (m, 5H), 3.94 (s, 3H), 3.93 (s, 3H), 3.83 (dd, 1H, *J* = 11.2, 4.8 Hz), 3.75 (dd, 1H, *J* = 3.2, 10.8 Hz), 2.63 (td, 1H, *J* = 6, 12.8 Hz), 2.41 (m, 1H), 2.33 (m, 2H), 0.914 (s, 9H), 0.911 (s, 9H), 0.11, (s, 6H), 0.08 (s, 6H). ^{13}C NMR ($CDCl_3$): δ 154.3, 153.8, 153.0, 152.8, 150.1, 148.4, 139.7, 137.6, 126.7, 120.8, 87.9, 84.1, 72.3, 66.6, 63.0, 56.7, 56.6, 41,2, 30.7, 26.1, 25.9, 18.6, 18.2, -4.5, -4.6, -5.2, -5.3. IR (neat): 2953, 2928, 2856, 1749, 1591, 1523, 1470, 1253, 1219, 1064, 834, 776 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{36}H_{57}N_6O_{10}Si_2$ 789.3669, found 789.3661.



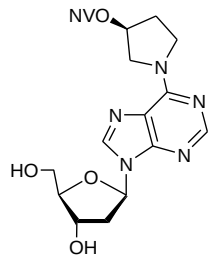
63b

Synthesis of 64a. Triethylamine trihydrofluoride (0.220 mg, 221 µL, 1.37 mmol) was added to a solution of levulinyl protected **63a** (310 mg, 0.47 mmol) in THF (3.0 mL). After stirring at room temperature for 17 h, TLC (5% MeOH in DCM) showed that a small amount of starting material was still present. An additional equivalent of desilylating reagent was added and the reaction stirred for 6 additional h at which time no starting material remained. The reaction was evaporated to dryness in vacuo, and 140 mg of a white foam-like product (71% yield) was collected after purification by flash



64a

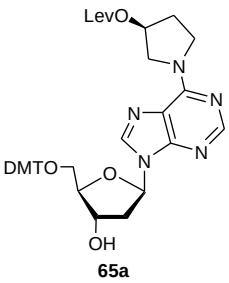
chromatography (2% MeOH in DCM to 5% MeOH in DCM). R_f (5% MeOH in DCM) = 0.3. ^{1}H NMR (MeOH-d_4): δ 8.23 (s, 1H), 8.18 (s, 1H), 6.42 (t, 1H, J = 6.8 Hz), 5.44 (m, 1H), 4.88 (m, 4H), 4.56 (d, 1H, J = 3.6 Hz), 4.10-4.50 (1H), 4.05 (d, 1H, J = 2.4 Hz), 3.87-4.02 (1H), 3.84 (dd, 1H, J = 11.2, 2.6 Hz), 3.73 (dd, 1H, J = 12.4, 2.9 Hz), 2.70-2.86 (3H), 2.52 (t, 2H, J = 6.4 Hz), 2.37 (m, 1H), 2.24 (m, 2H), 2.06 (s, 3H). ^{13}C NMR (MeOD-d_4): δ 209.4, 174.0, 154.2, 153.2, 150.4, 140.6, 121.8, 89.8, 87.2, 73.1, 63.6, 42.6, 38.6, 32.9, 31.1, 29.5, 29.0. IR (neat): 3263, 2929, 1713, 1593, 1472, 1204, 1158. 1095 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{19}H_{26}N_5O_6$ 410.1878, found 420.1887.



64b

Synthesis of 64b. Triethylamine trihydrofluoride (208 mg, 209 µL, 1.28 mmol) was added to a solution of NVOC protected **63b** (202 mg, 0.26 mmol) in THF (2.7 mL) and the mixture was stirred overnight at room temperature. After evaporating the solvent in vacuo the residue was dissolved in DCM and purified by flash chromatography (18 × 2.2 cm) (2% methanol in DCM to 4% methanol in DCM) to yield the desired product as a white foam (131 mg, 89%). R_f (4% methanol in DCM) = 0.15. ^{1}H NMR ($CDCl_3$): δ 8.29 (s, 1H), 7.77 (s, 1H), 7.72 (s, 1H), 7.01 (s, 1H), 6.86 (d, 1H, J = 8 Hz), 6.31 (dd, 1H, J = 5.6, 8 Hz), 5.59 (s, 2H), 5.45 (d, 1H, J = 16 Hz), 4.8 (d, 1H, J = 4 Hz), 4.57 (m, 1H), 3.96-4.34 (m, 4H), 3.95 (s, 3H), 3.94 (s, 3H), 3.79 (t, 1H, J = 8 Hz), 3.14 (m, 1H), 2.63 (s, 1H), 2.28 (d, 1H, J = 5.2 Hz), 2.25 (d, 1H, J = 5.5 Hz). ^{13}C NMR ($CDCl_3$): δ 154.3, 153.8, 153.4, 152.1, 149, 148.5, 139.8, 139.1, 126.5, 122.2, 110.2, 108.3, 89.9, 87.9, 77.4, 73.8, 66.6, 63.7, 56.6, 56.5, 54.7, 53.2, 46.5, 45.4, 40.6, 32.3, 30.2, 29.7, 8.9. IR (neat): 3332 (bd), 2934, 2270, 1750, 1596, 1522, 1471, 1387, 1332, 1277, 1220, 1064, 987, 792,

668 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{24}H_{29}N_6O_{10}$ 561.1940, found 561.1944.

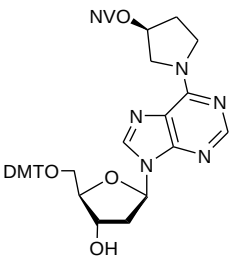


65a

Synthesis of 65a. A pyridine (2 mL) solution of DMT-Cl (73 mg, 0.19 mmol) and azeotropically dried **64a** (75 mg, 0.19 mmol) was stirred overnight, at which time the reaction was quenched with MeOH, and evaporated to dryness in vacuo. The dry mass was dissolved in EtOAc (20 mL), and washed with saturated sodium bicarbonate (15 mL) and dried over sodium sulfate. A flash silica column was prepared using 1% Et_3N in hexanes that was washed with 5 column volumes of 1% MeOH in EtOAc prior to loading the mixture. A solvent combination of 1% MeOH in EtOAc that was increased to 3% MeOH in EtOAc was used to obtain 93 mg of product (68%). R_f (5% MeOH in DCM) = 0.4. ^{1}H NMR ($CDCl_3$): δ 8.29 (s, 1H), 7.86 (s, 1H), 7.38 (d, 2H, *J* = 8.6 Hz), 7.14-7.34 (m, 7H), 6.78 (d, 4H, *J* = 8.6 Hz), 6.45 (t, 1H, *J* = 6.0 Hz), 5.48 (s, 1H), 4.64 (m, 1H), 4.15-4.55 (m, 2H), 4.12 (m, 1H), 3.72-4.09 (m, 2H), 3.78 (s, 3H), 3.75 (s, 3H), 3.36 (m, 2H), 3.05 (bd s, 1H), 2.68-2.85 (m, 3H), 2.44-2.59 (m, 3H), 2.21 (m, 2H), 2.15 (s, 3H). ^{13}C NMR ($CDCl_3$): δ 206.4, 172.3, 158.5, 153.0, 152.6, 149.8, 144.5, 137.2, 135.7, 130.0, 128.0, 127.9, 126.9, 120.7, 113.1, 86.5, 85.8, 83.9, 72.6, 63.8, 55.2, 54.0, 53.1, 47.0, 45.8, 40.3, 37.8, 30.7, 24.2, 29.8, 28.0. IR (neat): 3300 (bd), 2932, 2836, 1720, 1592, 1508, 1443 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{40}H_{44}N_5O_8$ 722.3184, found 722.3188.

Synthesis of 65b. A pyridine (3 mL) solution of DMT-Cl (108 mg, 0.26 mmol) and azeotropically dried (pyridine) **64b** (131 mg, 0.23 mmol) was stirred overnight, at which time the reaction was diluted with 50 mL EtOAc. The diluted contents were washed with

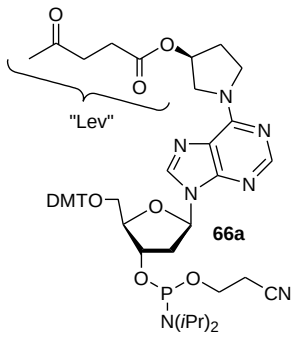
saturated ammonium chloride (2 × 50 mL), followed by washing with brine (50 mL). The aqueous layers were washed with EtOAc (50 mL) and the combined organic layers were dried over sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in DCM containing a few drops of Et_3N. A silica column packed with 1% Et_3N in DCM was used for separation. The crude product was purified by flash chromatography (60% EtOAc in DCM with a few drops of Et_3N to 3% methanol with 90% EtOAc in DCM with a few drops of Et_3N) to yield the desired product as a faint yellow foam (125 g, 64%). R_f (3% methanol and 90% EtOAc in DCM) = 0.3. ^{1}H NMR ($CDCl_3$): δ 8.30 (s, 1H), 7.88 (s, 1H), 7.72 (s, 1H), 7.41 (d, 2H, J = 5.2 Hz), 7.20-7.38 (m, 7H), 7.01 (s, 1H), 6.81 (d, 4H, J = 12 Hz), 6.45 (t, 1H, J = 6.4 Hz), 5.58 (s, 2H), 5.45 (m, 1H), 4.64 (m, 1H), 4.03-4.58 (m, 5H), 3.94 (s, 3H), 3.91 (s, 3H), 3.77 (s, 6H), 3.38 (m, 2H), 2.76 (m, 1H), 2.51 (m, 1H), 2.31 (bd s, 3H). ^{13}C NMR ($CDCl_3$): δ 158.6, 154.3, 153.9, 153.1, 152.8, 150.0, 148.5, 144.7, 139.8, 137.5, 135.8, 131.1, 130.1, 128.9, 128.2, 128.0, 127.0, 126.6, 120.9, 113.3, 109.9, 108.3, 86.7, 85.9, 84.1, 72.8, 66.6, 63.9, 56.7, 56.5, 55.3, 46.2, 40.4, 38.6, 32.1, 30.5, 29.8, 26.1, 23.9, 23.1, 14.3, 11.1, 8.9. IR (neat): 3314 (bd), 2926, 2853, 1749, 1593, 1508, 1250, 1219, 1064, 909, 729 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{45}H_{47}N_6O_{12}$ 863.3246, found 863.3236.



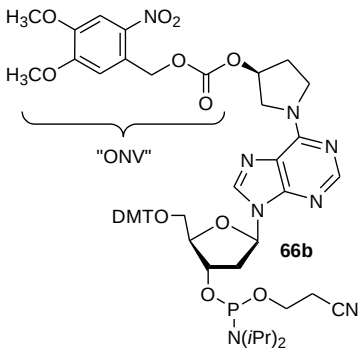
65b

Synthesis of 66a. DIPEA (85 mg, 0.66 mmol) was added to **65a** (79.2 mg, 0.11 mmol) that was azeotropically dried with pyridine followed by 0.2 mL of freshly distilled DCM. The solution was cooled in an ice bath and 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (23 mg, 0.11 mmol) was added. After stirring the reaction mixture for 1 h, the reaction mixture was diluted with EtOAc (20 mL), washed

with saturated sodium bicarbonate (15 mL), and then dried over sodium sulfate. After the organic layer was concentrated in vacuo, the compound was purified by flash chromatography (15 × 1.5 cm, packed using 0.5% Et_3N in hexanes and then washed with 5 column volumes of EtOAc) using EtOAc as eluent to isolate **66a** (79.6 mg, 78%). R_f (100% EtOAc) = 0.5. ^{1}H NMR ($CDCl_3$): δ 8.31 (s, 1H), 7.90 (d , 1H, J = 7.7 Hz), 7.35-7.46 (m, 2H), 7.13-7.32 (m, 7H), 6.72 (m, 4H), 6.44 (t, 1H, J = 5.1 Hz), 5.47 (s, 1H), 4.70 (m, 1H), 4.28 (m, 1H), 3.84-4.57 (m, 3H), 3.76 (s, 6H), 3.64 (m, 1H), 3.57 (m, 2H), 3.37 (m, 2H), 3.29 (m, 2H), 2.79 (m, 1H), 2.73 (d, 1H, J = 6.4 Hz), 2.69 (d, 1H, J = 6.4 Hz), 2.63 (m, 1H), 2.55 (m, 2H), 2.12 (m, 2H), 2.14 (s, 3H), 1.07-1.19 (m, 12H). ^{31}P NMR ($CDCl_3$): 149.06, 148.95. IR (neat): 2960, 1720, 1593, 1509, 1469, 1250, 1178, 1077 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{49}H_{60}N_7O_9P$ 922.4263, found 922.4265.

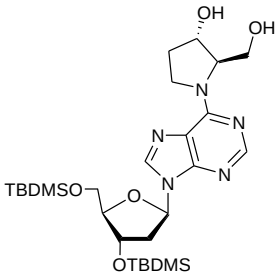


Synthesis of 66b. DIPEA (94 mg, 0.73 mmol) was added to **65b** (118 mg, 0.13 mmol) that was azeotropically dried with pyridine, followed by 1.5 mL of freshly distilled DCM. The solution was cooled in an ice bath and 2-cyanoethyl N,N-diisopropylchlorophosphoramidite (42.6 mg, 0.11 mmol) was added. After stirring the reaction mixture for 1 h, the reaction mixture was diluted with EtOAc (30 mL), washed with saturated sodium bicarbonate (20 mL), and then dried over sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in 60% ethyl in hexanes. A silica column was packed with 1% Et_3N in hexanes and then washed with 3 column volumes volume of 60% EtOAc in



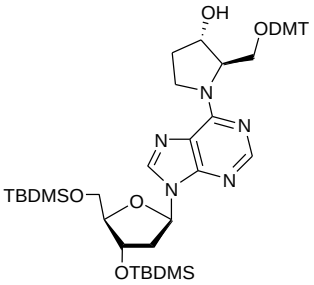
hexanes. The above solution was loaded onto the column and separated (60% EtOAc in hexanes to 90% EtOAc in hexanes) to yield the desired product (108 mg, 75%). R_f (90% EtOAc in hexanes) = 0.6. ^{1}H NMR ($CDCl_3$): δ 8.31 (s, 1H), 7.92 (d, 1H, J = 6.2 Hz), 7.74 (s, 1H), 7.39 (m, 2H, J = 11.2 Hz), 7.15-7.33 (m, 7H), 7.01 (s, 1H), 6.72-6.84 (m, 4H), 6.45 (t, 1H, J = 8 Hz), 5.59 (s, 2H), 5.45 (bd s, 1H), 4.72 (m, 1H), 4.00-4.60 (m, 5H), 3.95 (s, 3H), 3.92 (s, 3H), 3.77 (s, 6H), 3.26-3.71 (m, 5H), 2.82 (m, 1H), 2.75 (t, 1H, J = 4 Hz), 2.52-2.67 (m, 2H), 2.44 (t, 1H, J = 4 Hz), 2.34 (bd s, 2H), 1.08-1.57 (m, 12H). ^{31}P NMR ($CDCl_3$): δ 148.9, 148.7. IR (neat): 2925, 2852, 1749, 1592, 1509, 1467, 1249, 1065, 1033, 977, 793 cm^{-1}. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{54}H_{63}N_8O_{13}P$ 1063.0976, found 1085.3 ($M+Na^+$).

Synthesis of 109. A DMF (130 mL) solution of (2*R*,3*S*)-2-(hydroxymethyl)-3-hydroxypyrrolidine hydrochloride[137] (**108**, 2.02 g, 13.1 mmol) and Cs_2CO_3 (17.1 g, 52.6 mmol) were stirred at room temperatures for 30 min, at which time **9** (9.4 g, 15.7 mmol) was added. The reaction was stirred overnight at room temperature followed by dilution with 400 mL EtOAc. The diluted contents were washed with saturated sodium bicarbonate (300 ml), followed by washing with brine (2 × 200 mL). The combined aqueous layers were washed with EtOAc (200 mL) and the combined organic layers were dried over sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in DCM. It was purified by flash chromatography (30-80% EtOAc in DCM to 3% methanol and 7% DCM in EtOAc) to yield the desired product as a white solid (3.54 g, 46%). R_f (90% EtOAc in DCM) = 0.2. ^{1}H NMR ($CDCl_3$): δ 8.27 (s, 1H), 8.04 (s, 1H), 6.42 (t, 1H, J = 8.4 Hz), 4.58 (bd s, 2H),



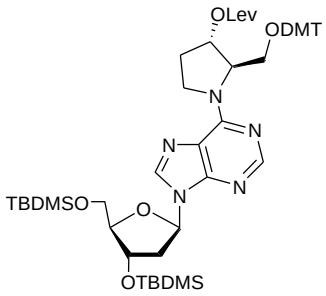
109

4.02-4.43 (m, 3H), 3.99 (dd, 1H, *J* = 4, 7.1 Hz), 3.85 (dd, 1H, *J* = 4.8, 11.9 Hz), 3.61-3.80 (m, 3H), 2.51-2.74 (m, 2H), 2.41 (m, 1H), 2.01-2.26 (m, 2H), 1.87 (bd s, 1H), 0.89 (s, 18H), 0.08 (s, 6H), 0.06 (s, 6H). ^{13}C NMR ($CDCl_3$): δ 154.0, 152.5, 150.1, 137.8, 120.6, 88.0, 84.4, 84.4, 73.9, 72.1, 70.5, 65.5, 62.9, 41.3, 26.8, 26.2, 25.99, 25.97, 18.6, 18.2, -4.4, -4.5, -5.1, -5.2. IR (neat): 3331 (bd), 2950, 2929, 2857, 1591, 1471, 1254, 1099, 1071, 836, 778 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ $C_{27}H_{50}N_5O_5Si_2$ calculated for 580.3345, found 580.3335.



Synthesis of DMT-protected 109. After azeotropically drying with pyridine, **109** (1.01 g, 1.7 mmol) was dissolved in a mixture of DMT-Cl (859.6 mg, 2.09 mmol) and, dimethylaminopyridine (43 mg, 0.348) mmol) in pyridine (20 mL) at 0°C. The reaction mixture was allowed to slowly warm to room temperatures and then stirred overnight, at which time it was diluted with 150 mL of EtOAc. The diluted contents were washed with saturated ammonium chloride (2 × 200 mL) and then with brine (100 mL). The combined aqueous layers were washed once with EtOAc (100 mL) and the combined organic layers were dried with sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in DCM containing a few drops of Et_3N. A flash chromatography on a column packed with 1% Et_3N in DCM (30% EtOAc in DCM to 70% EtOAc in DCM, along with a few drops of Et_3N) yielded the desired product as a faint yellow foam (1.19 g, 72%). R_f (30% EtOAc in DCM) = 0.2. ^{1}H NMR ($CDCl_3$): δ 8.29 (s, 1H), 7.93 (s, 1H), 7.38 (d, 2H, *J* = 8.2 Hz), 7.10-7.29 (m, 7H), 6.70 (m, 4H), 6.43 (t, 1H, *J* = 6.8 Hz), 4.59 (m, 2H), 4.05-4.21 (m, 1H), 4.01 (dd, 1H, *J* = 4, 7.2 Hz), 3.81 (m, 1H), 3.75-3.80 (m, 4H), 3.74 (m, 5H), 3.58

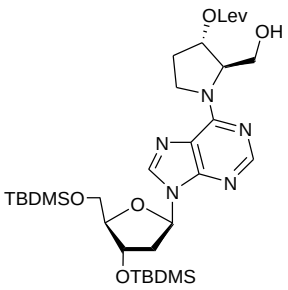
(dd, 1H, *J* = 3.6, 9.2 Hz), 3.27 (dd, 1H, *J* = 9.2, 6.8 Hz), 2.59 (td, 1H, *J* = 7.2, 12.4 Hz), 2.39 (m, 1H), 2.33 (m, 1H), 2.00 (bd s, 1H), 1.75 (bd s, 1H), 0.92 (s, 9H), 0.90 (s, 9H), 0.11 (s, 6H), 0.08 (s, 3H), 0.07 (s, 3H). ^{13}C NMR ($CDCl_3$): δ 158.35, 158.33, 153.2, 152.4, 144.8, 137.1, 136.7, 136.2, 136.0, 130.0, 129.9, 128.2, 127.6, 126.6, 120.6, 112.96, 112.92, 87.7, 86.4, 84.1, 77.2, 74.5, 72.1, 67.0, 62.9, 60.3, 55.1, 46.9, 41.0, 32.3, 25.9, 25.7, 18.4, 18.0, -4.67, -4.79, -5.3, -5.4. IR (neat): 3301 (bd), 3034, 2952, 2856, 1737, 1589, 1508, 1469, 1250, 1093, 1033, 834, 779 cm^{-1}. HRMS (ESI/APCI) *m/z* [M + H]$^+$ calculated for $C_{48}H_{68}N_5O_7Si_2$ 882.4652, found 882.4635.



110

Synthesis of 110. 1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide hydrochloride (1.15 g, 6 mmol) and DMAP (655 mg, 5.26 mmol) in DMF (10 mL) were added to DMT protected **109** (715 mg, 0.75 mmol). After stirring for 10 min at room temperature, levulinic acid (608 mg, 533 µL, 5.25 mmol) was added to the reaction mixture. The reaction mixture was stirred for 2 h followed by dilution with 70 mL EtOAc. The solution was washed with saturated sodium bicarbonate (1 × 70 mL) and then with brine (2 × 50 mL). The combined aqueous layers were washed once with EtOAc (50 mL) and the combined organic layers were dried with sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in DCM containing a few drops of Et_3N. It was purified by flash chromatography (10% EtOAc in DCM with few drops of Et_3N to 35% EtOAc in DCM with a few of drops of Et_3N) to yield the desired product as a white foam (714 mg, 89%). R_f (30% EtOAc in DCM) = 0.4. ^{1}H NMR ($CDCl_3$) at 39°C: δ 8.27 (bd s, 1H), 7.91 (bd s, 1H), 7.15-7.26 (m, 9H), 6.66-6.71 (m, 4H), 6.43 (bd s, 1H), 5.51 (bd s, 1H), 4.62 (bd s, 1H), 4.03-5.21 (m,

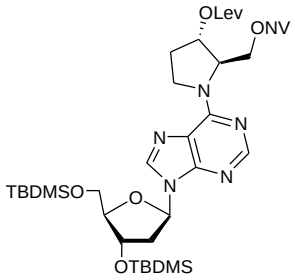
1H), 4.13 (m, 1H), 4.02 (m, 2H), 3.82 (m, 1H), 3.79 (m, 1H), 3.75 (s, 3H), 3.74 (s, 3H), 3.57 (m, 2H), 2.60-2.78 (m, 3H), 2.47-2.59 (m, 3H), 2.43 (m, 1H), 2.13 (m, 4H), 0.94 (s, 9H), 0.91 (s, 9H), 0.11 (s, 6H), 0.08 (s, 6H). ^{13}C NMR ($CDCl_3$) at 39°C: δ 206.3, 172.3, 158.67, 158.64, 153.4, 152.7, 150.7, 145.1, 137.4, 136.4, 136.2, 130.29, 130.20, 128.4, 127.8, 126.8, 121.2, 113.2, 113.2, 88.0, 86.9, 84.3, 77.2, 72.5, 64.6, 63.2, 55.3, 41.2, 38.1, 29.9, 28.5, 26.3, 26.0, 18.6, 18.2, -4.4, -4.5, -5.1, -5.2. IR (neat): 2961, 2929, 2856, 1725, 1720, 1588, 1509, 1468, 1251, 1177, 1088, 1033, 835 cm^{-1}. HRMS (ESI/APCI) *m/z* [M + H]$^+$ calculated for $C_{53}H_{74}N_5O_9Si_2$ 980.5020, found 980.5028.



Synthesis of DMT deprotected 110. Acetic acid, water and tetrahydrofuran (3:1:1, 50 mL) were added to **110** (715 mg, 0.67 mmol). After stirring for 22 min at room temperature, methanol (2 mL) was added and the solution was stirred for an additional 8 min, at which time 300 mL of EtOAc was added. The solution was washed with water (3 × 100 mL). The combined aqueous layers were washed with fresh EtOAc (200 mL). The combined organic layers were carefully washed with saturated sodium bicarbonate (400 mL) and then dried over sodium sulfate. After evaporating the solvent in vacuo, the product mixture was dissolved in 30% EtOAc in hexanes and purified by flash chromatography (50% EtOAc in hexanes to 80% EtOAc in hexanes) to yield the desired product as a white foam (403 mg, 89%). R_f (70% EtOAc in hexanes) = 0.3. ^{1}H NMR ($CDCl_3$): δ 8.32 (s, 1H), 8.06 (s, 1H), 6.44 (t, 1H, *J* = 6.8 Hz), 5.22 (bd s, 1H), 4.75 (bd s, 1H), 4.60 (td, 1H, *J* = 5.2, 3.2 Hz), 4.23 (m, 2H), 3.99 (dt, 1H, *J* = 3.2, 7.1 Hz), 3.93 (m, 1H), 3.84 (dd, 1H, *J* = 4.3, 10.6 Hz), 3.75 (dd, 1H, *J* = 3.1, 10.6 Hz), 3.71 (dd, 1H, *J* = 7.1, 11.24 Hz), 2.69 (m, 2H), 2.61 (td, 1H, *J* = 6.4, 12.4 Hz), 2.52

(m, 2H), 2.42 (m, 1H), 2.14-2.39 (m, 3H), 2.12 (s, 3H), 0.92 (s, 18H), 0.10 (s, 6H), 0.08 (s, 6H). ^{13}C NMR ($CDCl_3$): δ 206.7, 206.2, 172.1, 153.7, 152.3, 150.1, 137.8, 87.8, 84.7, 77.2, 71.9, 67.0, 65.3, 62.7, 41.1, 37.7, 30.85, 30.81, 29.79, 29.73, 29.6, 28.0, 25.9, 25.7, 18.3, 17.9, -4.6, -4.8, -5.4, -5.5. IR (neat): 3363 (bd), 2953, 2928, 2856, 1719, 1589, 1470, 1253, 1157, 1110, 1071, 1029, 836, 778 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{32}H_{56}N_5O_7Si_2$ 678.3726, found 678.3729.

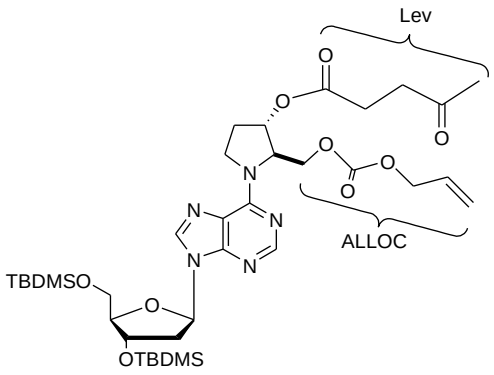
Synthesis of 111a. Nitroveratryloxycarbonyl chloroformate[164] (326 mg, 1.18 mmol), DMAP (277 mg, 2.23 mmol), and DCM (6 mL) were added to azeotropically dried (pyridine) DMT deprotected **110** (310 mg, 0.447 mmol)**.** After stirring at room temperature for 2 h, the reaction mixture was diluted with 100 mL DCM. The diluted solution was washed with sodium bicarbonate (50 mL) and brine (50 mL). The aqueous layers were washed with DCM (50 mL). The combined organic layers were dried over sodium sulfate. After evaporating the solvent in vacuo, the crude was dissolved in 30% EtOAc in hexanes and purified by flash chromatography (30-70% EtOAc in hexanes) to yield the desired product as a yellowish foam (349 mg, 83%). R_f (50% EtOAc in hexanes) = 0.3. ^{1}H NMR ($CDCl_3$): δ 8.35 (s, 1H), 8.01 (s, 1H), 7.71 (s, 1H), 7.02 (s, 1H), 6.43 (t, 1H, *J* = 8 Hz), 5.55 (s, 2H), 5.42 (bd s, 1H), 4.59 (m, 1H), 4.55 (m, 2H), 4.17 (m, 1H), 3.98 (m, 2H), 3.89-3.97 (m, 7H), 3.81 (dd, 1H, *J* = 4.4, 11.2 Hz), 3.74 (dd, 1H, *J* = 4.3, 11.2 Hz), 2.70 (m, 2H), 2.63 (m, 1H), 2.53 (m, 2H), 2.41 (m, 2H), 2.18 (m, 1H), 2.12 (s, 3H), 0.90 (s, 9H), 0.89 (s, 9H), 0.09 (s, 6H), 0.07 (s, 3H), 0.06 (s, 3H). ^{13}C NMR ($CDCl_3$): δ 206.2, 172.0, 154.4, 153.7, 153.0, 152.5, 150.2, 148.3, 139.6, 137.9, 126.6, 120.7, 109.8, 108.1, 87.8, 84.0,72.0, 66.4, 63.0,



111a

62.8, 56.4, 56.3, 47.3, 40.9, 37.7, 29.9, 29.6, 29.6, 28.0, 25.9, 25.7, 18.3, 17.9, 18.3, 17.9, 15.3, -4.7, -4.8, -5.4, -5.5. IR (neat): 2953, 2928, 2856, 1738, 1587, 1524, 1466, 1278, 1253, 1221, 1065, 837, 780 cm-^{1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{43}H_{68}N_6O_{13}Si_2$ 917.4143, found 917.4128.

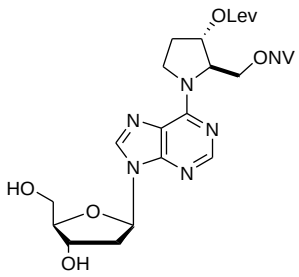
Synthesis of 111b. Allyl chloroformate (107 μL, 1 mmol), DMAP (372 mg, 3 mmol), and DCM (8 mL) were added to azeotropically dried (pyridine) DMT deprotected **110** (403mg, 0.59 mmol)**.** After stirring at room temperature for 2 h, the reaction mixture was diluted with 100 mL DCM. The diluted solution was washed with sodium bicarbonate (40 mL) and brine (40 mL). The aqueous layers were washed with DCM (30 mL). The combined organic layers were dried over sodium sulfate. After evaporating the solvent in vacuo, the crude was dissolved in DCM and purified by flash chromatography (20% EtOAc in DCM) to yield the desired product (354 mg, 77%). R_f (50% EtOAc in DCM) = 0.3. ^{1}H NMR ($CDCl_3$) δ 8.36 (s, 1H), 8.02 (s, 1H), 6.45 (t, *J* = 6.5 Hz, 1H), 5.90 (ddt, *J* = 16.4, 11.0, 5.8 Hz, 1H), 5.46 – 5.15 (m, 3H), 4.55 (dd, *J* = 21.9, 4.5 Hz, 5H), 4.00 (q, *J* = 3.7 Hz, 2H), 3.79 (ddd, *J* = 23.5, 11.1, 3.9 Hz, 3H), 2.84 – 2.29 (m, 8H), 2.26 – 2.07 (m, 4H), 0.96 – 0.80 (m, 18H), 0.21 (m, 12H). ESI (TOF) *m/z* $[M+H^+]$ calculated for $C_{36}H_{59}N_5O_9Si_2$ 762.0, found 762.1.



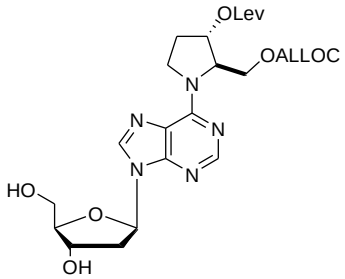
111b

Synthesis of 112a. A solution of triethylamine trihydrofluoride (291 mg, 293 μL, 1.8 mmol) and **111a** (339 mg, 0.36 mmol) in THF (4 mL) were stirred overnight at room temperature. After evaporating the solvent in vacuo the residue was dissolved in

dichloromethane and purified by flash chromatography (2-4% methanol in DCM) to yield the desired product as a white foam (271 mg, 93%). R_f (4% methanol in DCM) = 0.15. ^{1}H NMR ($CDCl_3$): δ 8.31 (s, 1H), 7.78 (s, 1H), 7.70 (s, 1H), 7.01 (s, 1H), 6.62 (d, 1H, *J* = 12 Hz), 6.30 (dd, 1H, *J* = 6, 8.8 Hz), 5.54 (s, 2H), 5.42 (m, 1H), 4.65-4.86 (m, 2H), 4.55 (m, 2H), 4.19 (s, 1H), 3.86-4.01 (m, 9H), 3.76 (m, 1H), 3.10 (m, 1H), 2.72 (m, 2H), 2.37-2.59 (m, 3H), 2.17-2.33 (m, 3H), 2.13 (s, 3H). ^{13}C NMR ($CDCl_3$): δ 206.3, 172.1, 154.4, 153.6, 153.4, 151.8, 149.3, 148.3, 139.6, 139.4, 129.4, 126.4, 110.0, 108.2, 89.6, 87.7, 77.6, 73.5, 66.5, 63.5, 63.2, 56.4, 56.3, 53.3, 40.4, 37.7, 29.7, 29.6, 27.9. IR (neat): 3341 (bd), 2927, 1717, 1588, 1523, 1468, 1330, 1277, 1220, 1158, 1096, 1065 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{31}H_{40}N_6O_{13}$ 689.2413, found 689.2445.



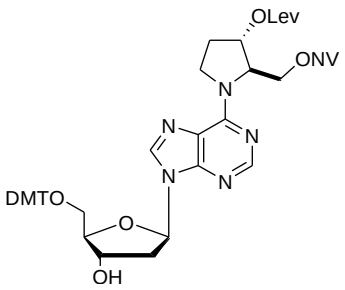
112a



112b

Synthesis of 112b. A solution of triethylamine trihydrofluoride (367 μL, 2.26 mmol) and **111b** (354 mg, 0.46 mmol) in THF (4 mL) were stirred overnight at room temperature. After evaporating the solvent in vacuo the residue was dissolved in dichloromethane and purified by flash chromatography (90% ethyl acetate in DCM to remove **112b**, then 5% methanol and 4% DCM in ethyl acetate to elute **112b**) to yield the desired product (215 mg, 87%). R_f (5% methanol and 4% DCM in ethyl acetate) = 0.2. ^{1}H NMR ($CDCl_3$): δ 8.29 (d, *J* = 6.2 Hz, 1H), 7.79 (s, 1H), 6.70 (d, *J* = 11.0 Hz, 1H), 6.30 (dd, *J* = 9.5, 5.5 Hz, 1H), 5.88 (ddt, *J* = 16.4, 10.4, 5.9 Hz, 1H), 5.39 (broad s, 1H), 5.35 – 5.20 (m, 2H), 4.77 (d, *J* = 5.0 Hz, 1H), 4.74 – 4.58 (broad s, 1H), 4.57 (d, *J* = 5.9 Hz, 3H), 4.47 (dd, *J* = 11.2, 3.2 Hz,

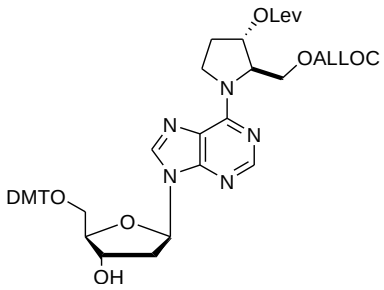
1H), 4.19 (s, 1H), 4.00 – 3.83 (m, 2H), 3.77 (t, *J* = 11.4 Hz, 1H), 3.10 (ddd, *J* = 14.1, 9.5, 5.0 Hz, 1H), 2.80 – 2.38 (m, 6H), 2.27 (dd, *J* = 13.4, 5.4 Hz, 1H), 2.17 (broad s, 1H), 2.13 (s, 3H), 1.83 (s, 1H). ^{13}C NMR ($CDCl_3$) δ 206.5, 172.3, 154.8, 153.6, 151.9, 139.5, 131.5, 119.3, 89.8, 87.8, 73.6, 68.8, 63.7, 63.5, 60.5, 40.7, 37.9, 29.8, 28.2, 21.1, 14.3. IR (neat): 3367 (broad), 2925, 1737, 1776, 1587, 1469, 1236, 1157, 1095, 941, 728, 646. ESI (TOF) *m/z* [M+H$^+$] calculated for $C_{24}H_{31}N_5O_9$ 533.5, found 534.1.



.Synthesis of DMT protected 112a. DMT-Cl (130 mg, 0.32 mmol), dimethylaminopyridine (1 mg, 0.01 mmol) and pyridine (4 mL) were added to azeotropically dried (pyridine) **112a** (186 mg, 0.26 mmol) at 0 °C. The reaction mixture was allowed to slowly warm to room temperature and stir overnight, at which time it was diluted with 50 mL of ethyl acetate. The solution was washed with saturated ammonium chloride (2 × 50 mL), followed by washing with brine (50 mL). The combined aqueous layers were washed with ethyl acetate (50 mL) and the combined organic layers were dried with sodium sulfate. After evaporating the solvent in vacuo, the crude mixture was dissolved in dichloromethane containing a few drops of Et_3N. The product was purified by flash chromatography (20-60% ethyl acetate in dichloromethane to 3% methanol in ethyl acetate, with a few drops of Et_3N) using a column packed with 1% Et_3N in DCM. The desired product was isolated as a faint yellow foam (182 mg, 64%). R_f (5% methanol in DCM) = 0.4. ^{1}H NMR ($CDCl_3$): δ 8.30 (s, 1H), 7.88 (s, 1H), 7.71 (s, 1H), 7.39 (d, 2H, *J* = 7.8 Hz), 7.19-7.33 (m, 7H), 7.03 (s, 1H), 6.79 (d, 4H, *J* = 9 Hz), 6.44 (t, 1H, *J* = 5.2 Hz), 5.55 (s, 2H), 5.43 (bd s, 1H), 4.63 (bd s, 1H), 4.25-4.82 (m, 5H), 4.07-4.19 (m, 2H), 3.95 (s, 3H), 3.94 (s, 3H), 3.77 (m, 7H), 3.40 (dd, 1H, *J* = 3.6,

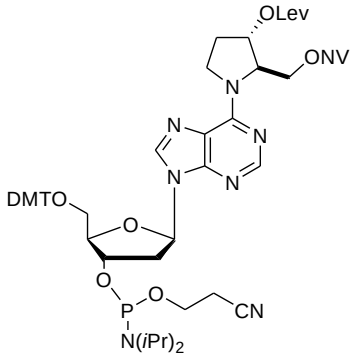
10.8 Hz), 3.38 (dd, 1H, *J* = 5.4, 10.8 Hz), 2.67-2.80 (m, 3H), 2.36-2.53 (m, 3H), 2.23 (m, 1H), 2.12 (s, 3H). ^{13}C NMR ($CDCl_3$): δ 206.2, 172.1, 158.5, 154.4, 153.7, 153.1, 152.5, 150.2, 148.3, 144.5, 139.6, 137.7, 135.7, 130.0, 128.0, 127.8, 126.9, 126.5, 120.8, 113.2, 109.9, 108.2, 86.6, 85.7, 83.9, 80.2, 77.7, 72.7, 66.4, 63.8, 63.1, 60.5, 56.5, 56.4, 55.2, 47.5, 40.1, 37.7, 30.6, 29.7, 28.0, 26.1. IR (neat): 2982-2939 (bd), 1734, 1587, 1509, 1372, 1236, 1176, 1157, 1043, 1064 cm^{-1}. HRMS (ESI/APCI) *m/z* $[M + H]^+$ calculated for $C_{52}H_{58}N_6O_{15}$ ($M+H^+$) 991.3720, found 991.3726.

Synthesis of DMT protected 112b. DMT-Cl (206 mg, 0.49 mmol), dimethylaminopyridine (DMAP, 2 mg) and pyridine (4.5 mL) were added to azeotropically dried (pyridine) **112a** (212 mg, 0.40 mmol) at 0 °C. The reaction mixture was allowed to slowly warm to room temperature and stir overnight. The solvent was removed in vacuo and the residue was redissolved in DCM containing a few drops of Et_3N. The product was purified by flash chromatography (70% ethyl acetate in dichloromethane to 2% methanol and 8% DCM in ethyl acetate, with a few drops of Et_3N) using a column packed with 1% Et_3N in DCM. The desired product was isolated as a faint yellow foam (236 mg, 71%). R_f (2% methanol and 8% DCM in ethyl acetate) = 0.5. ^{1}H NMR ($CDCl_3$) δ δ 8.30 (s, 1H), 7.87 (d, *J* = 3.3 Hz, 1H), 7.44 – 7.36 (m, 2H), 7.36 – 7.13 (m, 9H), 6.88 – 6.73 (m, 4H), 6.43 (t, *J* = 6.5 Hz, 1H), 5.89 (ddt, *J* = 16.4, 10.3, 5.8 Hz, 1H), 5.41 (s, 1H), 5.37 – 5.19 (m, 2H), 4.70 – 4.61 (m, 1H), 4.61 – 4.43 (m, 4H), 4.11 – 3.98 (m, 1H), 3.78 (s, 7H), 3.40 (qd, *J* = 10.1, 5.1 Hz, 2H), 2.81 – 2.60 (m, 3H), 2.60 – 2.37 (m, 4H), 2.19 (d, *J* = 23.2 Hz, 2H), 2.12 (s, 3H). IR (neat): 3367 (broad), 2918, 2849, 1742, 1588, 1508, 1466, 1249,

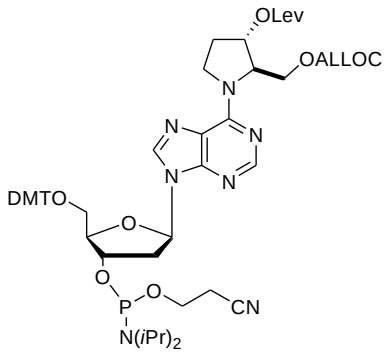


1177, 1033, 953, 829, 728, 703, 648, 583 cm^{-1}.ESI (TOF) *m/z* $[M+H^+]$ calculated for $C_{45}H_{49}N_5O_{11}$ 835.87, found 858.3 $[M+Na^+]$.

Synthesis of 113a. DMT protected **112a** (110 mg, 0.11 mmol) was azeotropically dried with pyridine. 2-Cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (42.2 mg, 44 μL, 0.19 mmol), diisopropylethylamine (120 mg, 161 μL, 0.9 mmol) and dichloromethane (3 mL) were added and the reaction mixture was stirred for 45 min at room temperature, at which time it was diluted with 60 mL of ethyl acetate. The solution was washed with saturated sodium bicarbonate (25 mL), followed by brine (25 mL), and then dried with sodium sulfate. After evaporating the solvent in vacuo, the mixture was dissolved in 60% ethyl acetate in hexanes and purified (60-90% ethyl acetate in hexanes) on a silica column packed with 1% Et_3N in hexanes, and washed with 3 column volumes of 60% ethyl acetate in hexanes. The yield of the desired product was 101.1 mg (78%). R_f (90% EtOAc in hexanes) = 0.6. ^1H NMR ($CDCl_3$): δ 8.31 (s, 1H), 7.91 (d, 1H, *J* = 7.1 Hz), 7.72 (s, 1H), 7.40 (m, 2H), 7.15-7.34 (m, 7H), 7.04 (s, 1H), 6.79 (m, 4H), 6.44 (t, 1H, *J* = 6.4 Hz), 5.55 (s, 2H), 5.42 (bd s, 1H), 4.57-4.71 (m, 4H), 4.29 (m, 1H), 3.94 (m, 7H), 3.77 (s, 6H), 3.58 (m, 2H), 3.38 (m, 3H), 3.29 (m, 3H), 2.63-2.84 (m, 4H), 2.36-2.63 (m, 4H), 2.21 (bd s, 1H), 2.11 (s, 3H), 1.05-1.20 (m, 12H). ^{31}P NMR ($CDCl_3$): δ 148.7, 148.5. IR (neat): 2965, 2932, 1736, 1586, 1623, 1508, 1465, 1364, 1329, 1278, 1246, 1221, 1178, 1156, 1065, 1032, 975, 823, 793, 727 cm^{-1}. HRMS (ESI) *m/z* $[M + H]^+$ calculated for $C_{60}H_{71}N_8O_{16}P$ 1191.2235, found 1191.3.



113a



113b

Synthesis of 113b. DMT protected **112b** (235 mg, 0.28 mmol) was azeotropically dried with pyridine. 2-Cyanoethyl-*N*,*N*-diisopropylchlorophosphoramidite (42.2 mg, 0.19 mmol), diisopropylethylamine (261 μL, 1.8 mmol) and dichloromethane (3 mL) were added and the reaction mixture was stirred for 45 min at room temperature, at which time it was diluted with 60 mL of ethyl acetate. The solution was washed with saturated sodium bicarbonate (40 mL), followed by brine (40 mL), and then dried with sodium sulfate. After evaporating the solvent in vacuo, the mixture was dissolved in 60% ethyl acetate in hexanes and purified (60-90% ethyl acetate in hexanes) on a silica column packed with 1% Et_3N in hexanes, and washed with 3 column volumes of 60% ethyl acetate in hexanes. The yield of the desired product was 203.1 mg (70%). R_f (90% EtOAc in hexanes) = 0.8. ^{1}H NMR δ 8.31 (s, 1H), 7.92 (d, *J* = 10.2 Hz, 1H), 7.40 (dd, *J* = 7.6, 3.3 Hz, 2H), 7.23 – 7.25 (m, 9H), 6.79 (dd, *J* = 8.6, 5.4 Hz, 4H), 6.45 (d, *J* = 4.9 Hz, 1H), 5.89 (m, 1H), 5.39 (s, 1H), 5.29 (m, 2H), 4.83 – 4.64 (m, 2H), 4.64 – 4.40 (m, 5H), 4.27 (dd, *J* = 6.1, 3.3 Hz, 1H), 3.72 – 3.78 (m, 9H), 3.66 – 3.41 (m, 4H), 3.44 – 3.27 (m, 2H), 2.87 – 2.44 (m, 10H), 2.17 (s, 2H), 2.13 (s, 3H), 1.42 – 1.16 (m, 12H). ^{31}P NMR ($CDCl_3$) δ 148.7, 148.6. IR (neat) 2965, 2929, 1743, 1719, 1623, 1587, 1508, 1466, 1248, 1177, 908, 726, 647 cm^{-1}. ESI (TOF) *m/z* [M+H$^+$] calculated for $C_{54}H_{66}N_7O_{12}P$ 1036.1, found 1058.5 [M+Na$^+$].

General Procedure for Synthesis of Template Strand Containing 67, 103, and 136. Native nucleotides were coupled using standard synthesis cycles (25 s coupling, 5 s capping with acetic anhydride, 15 s oxidation with 0.02 M I_2 in THF/H_2O/Pyridine, 95 s

detritylation with 3% TCA in DCM) for all oligonucleotides except for the **127** and **128**, where double coupling was utilized. Oligonucleotides **77**, **79**, **82**, **83**, **84**, **127** and **128** were prepared on 1000 Å resin. All other oligonucleotides were synthesized using 500 Å resin. Modified phosphoramidites **66a**, and **b** were coupled using automated double coupling cycles (15 min coupling time) while **113a** and **b** were coupled manually, as described below. The synthesizer was programmed to pause when it reached the point where the manual coupling was utilized.

General Procedure for Manual Coupling. Phosphoramidite **113a** (13 mg, 11 μmol) or **113b** (15 mg, 15 μmol) was dissolved in 300 μL activator solution (5-ethylthio-1H-tetrazole in acetonitrile, 0.25 M). The solution was repeatedly passed through a column for 10 min containing the partially synthesized oligonucleotide on CPG resin that was fitted with two 1 mL plastic syringes. The resin was washed with dry acetonitrile (5 mL) and dried under vacuum for 10 min. For manual double coupling, this procedure was repeated using a fresh solution of phosphoramidite and activator prior to carrying out the remaining solid phase synthesis steps on the automated synthesizer.

General Procedure for Trityl Quantification. The trityl output collected from the DNA synthesis was diluted in a volumetric flask using 0.1 M pTsOH in acetonitrile. The synthetic yield of the corresponding step was calculated based on the absorbance of trityl cation measured at 495 nm (ε_{495} = 70000).

General Procedure for Photolysis of Resin Containing 67 (ONV protected) and 103. Photolysis of the resin was carried out in 6-8 mg portions in toluene for 2 h in a Pyrex tube using a Schoeffel 1000 W short-arc lamp (model LH151N/2, intensity 9

mW/cm^2 at 350 nm) fitted with an optical filter (Newport # CGA-375). The resin was stirred using a pipe cleaner (to prevent crushing the resin) during the photolysis. After the photolysis, the resin was washed sequentially with hexanes (3 × 3mL), DCM (3 × 3 mL), and methanol (10 × 3 mL) within the Pyrex tube. After drying on a rotovap, the resin was transferred to a DNA synthesis column and dried under vacuum before any further synthetic modification. Methanol wash and drying in the previous step was repeated until complete recovery of resin was achieved.

General Procedure for Delevulinylation of Levulinyl Protected 67 and 105. A solution of 0.25 M hydrazine in pyridine/acetic acid (3:2) was passed back and forth through the resin in a DNA synthesis column as described above for the manual coupling procedure, for 10 min (for ≤ 25 nucleotide long) or 25 min (for longer substrates). For longer treatment times, fresh hydrazine solution was used every 6 min. At the end of the treatment, the resin was washed sequentially with methanol (5 × 10 mL) and acetonitrile (5 × 10 mL) and then dried under vacuum before continuing the oligonucleotide synthesis.

General Procedure for ALLOC Removal from 136. ALLOC removal of the resin was carried out in 6-8 mg portions of resin in a Pyrex vial fitted with a pipe cleaner. After adding $Et_2NH•BH_3$ (14 μM) in DCM (0.5 mL) and stirring for 5 min (room temperature), $Pd(PPh_3)_4$ (0.2 μM) in DCM (0.5 mL) was added. After stirring for 10 min (room temperature), the resin as allowed to settle and the supernatant was removed. The resin was resuspended and washed with DCM (4 × 1 mL). After the DCM was removed as a supernatant, the resin was treated with $Et_2NH•BH_3$ and $Pd(PPh_3)_4$ as described

above. Following the removal of $Et_2NH.BH_3$ and $Pd(PPh_3)_4$ and DCM wash (4 × 1 mL), the resin was dried on a rotavap and transferred to a DNA synthesis column. The resin was then fitted with a glass luer lock syringe and successively washed with 0.2% trifluoroacetic acid in DCM (15 × 1 mL), DCM (15 × 1 mL), 5% DIPEA in DCM (15 × 1 mL), DCM (15 × 1 mL). In the modified washing procedure, this was followed by acetonitrile (10 × 1 mL) and methanol (10 × 1 mL) wash using a luer lock syringe. The resin in the DNA synthesis column was then dried under vacuum before any continuing solid phase synthesis.

Table 13. Solvent conditions for HPLC chromatogram for purification of 79 containing DOB ICL mimic 3

Time	Flow	%A (50 mM TEAA)	%B (ACN)
0.00	4.00	95.0	5.0
5.00	4.00	92.0	8.0
50.00	4.00	87.0	13.0
55.00	4.00	80.0	20.0
60.00	4.00	60.0	40.0
65.00	4.00	20.0	80.0
70.00	4.00	95.0	5.0
81.00	4.00	95.0	5.0
81.10	4.00	95.0	5.0
82.10	4.00	95.0	5.0

HPLC Purification of 79. Upon completion of solid phase synthesis (under "trityl on" mode for the final coupling), the oligonucleotide was deprotection and cleaved from the resin by ammonolysis (65 °C, overnight). After removal of ammonia and water, the crude oligonucleotide mixture was resuspended in 50 mM TEAA (100 μL), filtered (Whatman nylon membrane syringe filter, 0.22 μm) and subjected to HPLC analysis on a RP-C18 HPLC column with monitoring carried out at 260 nm. The peaks were analyzed using the following gradient conditions (Table 13). The relevant peak of interest was

collected from the LC, lyophilized to dryness, detritylated, and analyzed by MALDI-TOF MS.

LC/MS Characterization of Synthesized Cross-links. Cross-links (25 μM) were dissolved in H_2O and filtered (Whatman nylon membrane syringe filter, 0.22 μm) prior to analysis. The injection volume was 5 – 10 μL. An orthogonal ESI-Q-TOF mass spectrometer (Xevo-G2, Waters) was connected in-line with an ACQUITY H-class UPLC. Mass spectra were acquired in negative ion mode using an ESI-MS capillary voltage of 2.5 kV, a sample cone voltage of 35 V and an extraction cone voltage of 4 V The cone gas flow was set up to 0 L/h and desolvation gas flow was 400 L/h. Desolvation temperature and source temperature were set to 350 and 120 °C, respectively. The acquisition range as m/z 500 – 3500. The LC/MS system was operated by Mass Lynx software. Raw summed spectra were deconvoluted using the MaxEnt1 algorithm.

Table 14. Solvent conditions for LC/MS characterization of synthesized cross-links.

Time (min)	% A (15 mM TEA, 200 mM HFIP)	%B (50% MeOH, 50% mobile phase A)
0.0	90	10
2.0	90	10
30.0	55	45
31.0	10	90
33.0	10	90
33.1	90	10

LC was performed using a linear gradient containing mobile phase A (15 mM TEA, 200 mM HFIP) to mobile phase B (50% MeOH, 50% mobile phase A) on a ACQUITY UPLC®BEH C18 column (17 μM, 21 × 50 mm) using the following linear gradient (Table 14). The built in column heater of the UPLC system was set to 60 °C and mobile phase flow rate was 0.2 mL/min.

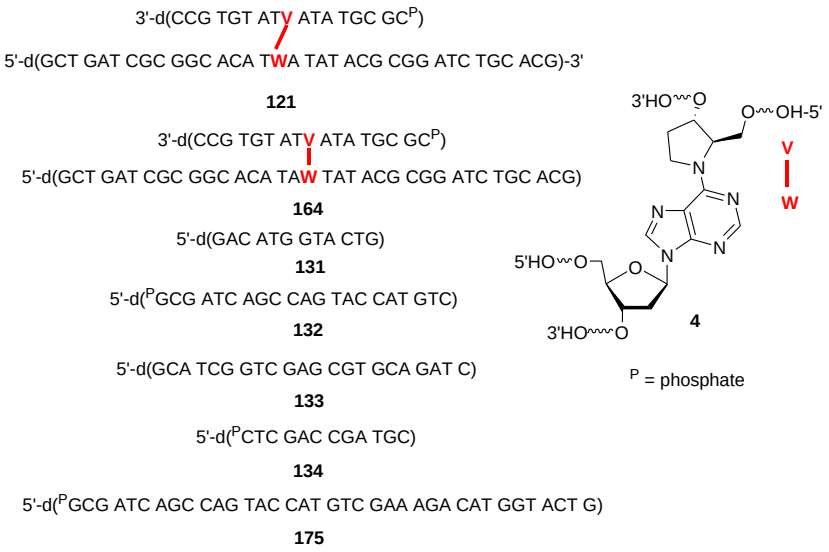
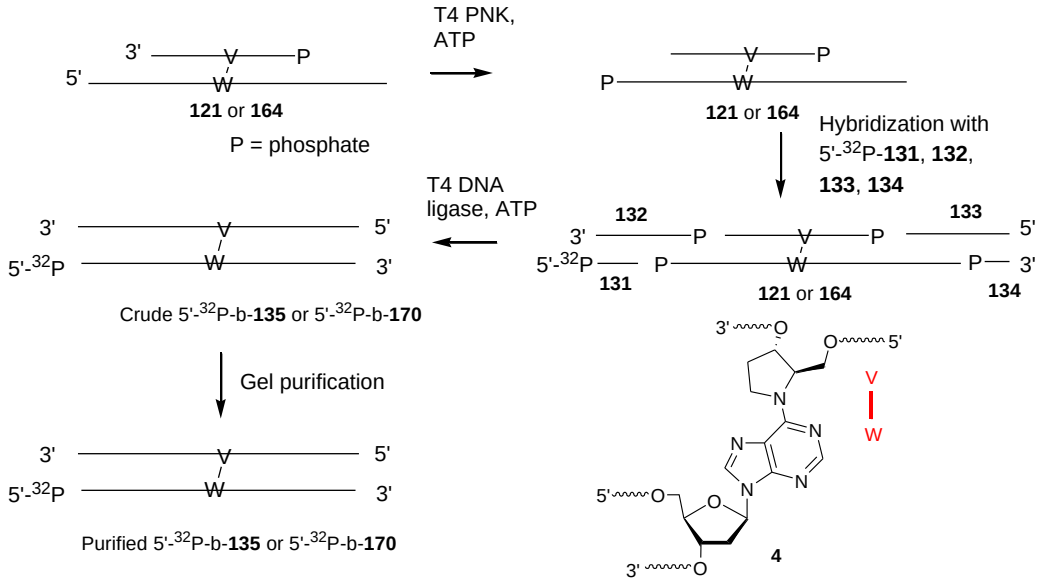


Figure 74. Oligonucleotides used for preparing **135**, **148**, and **170**.

General Procedure for Preparing 135. Scheme 45 describes the the general procedure for preparing **135** from **121** (Figure 74). Oligonucleotide **131** (80 pmol, Figure 74) was 5'-^{32}P-labeled in reaction (80 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 6% PEG 8000 solution (w/v), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 1 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **131** (A_{260}, ε =119700 L•mol^{-1}•cm^{-1}). Separately, **121** (40 pmol, Figure 74) was 5'-phosphorylated on the template strand in a 25 μL reaction containing 1 × T4 PNK buffer, ATP (80 pmol), and 30 U of T4 PNK at 37 °C for 1 h. Both phosphorylation reactions were stopped by heating at 65 °C for 30 min. The phosphorylated products were hybridized with 160 pmol each of **132**, **133**, and **134** (Figure 74)

Scheme 45. General Procedure for preparing 5'-^{32}P-b-135 and 5'-^{32}P-b-170



in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 µL, 90% formamide, 10 mM EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 5 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 5'-^{32}P-b-**135**. The yield (10-20%) of 5'-^{32}P-b-**135** was determined based on the specific activity of 5'-^{32}P-**131**.

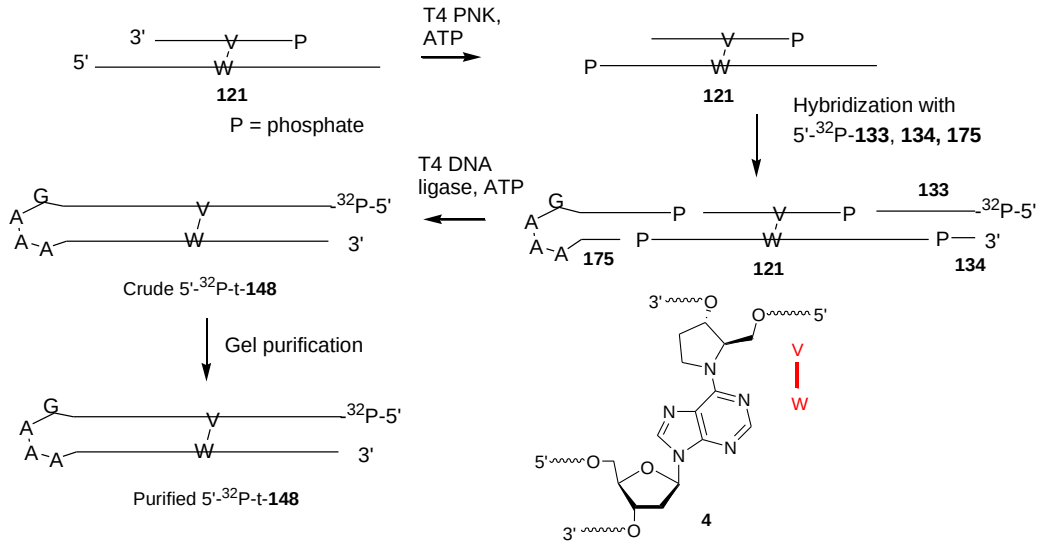
General Procedure for Preparing 5'-^{32}P-b-170. Scheme 45 describes the general procedure for preparing 5'-^{32}P-b-**170** from **164**. Oligonucleotide **131** (80 pmol, Figure 74) was 5'-^{32}P-labeled in reaction (80 µL) containing 1 × T4 PNK buffer (70 mM

TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 6% PEG 8000 solution (w/v), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 1 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **131** (A_{260}, ε =119700 L•mol^{-1}•cm^{-1}). Separately, **164** (40 pmol, Figure 74) was 5'-phosphorylated on the template strand in a 25 μL reaction containing 1 × T4 PNK buffer, ATP (80 pmol), and 30 U of T4 PNK at 37 °C for 1 h. Both phosphorylation reactions were stopped by heating at 65 °C for 30 min. The phosphorylated products were hybridized with 160 pmol each of **132**, **133**, and **134** (Figure 74) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 5 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 5'-^{32}P-b-**170**. The yield (10-20%) of 5'-^{32}P-b-**170** was determined based on the specific activity of 5'-^{32}P-**131**.

General Procedure for Synthesis of 5'-^{32}P-t-148. The preparation of 5'-^{32}P-t-**148** is described in Scheme 46. Oligonucleotide **133** (80 pmol, Figure 74) was 5'-^{32}P-labeled

in reaction (80 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 6% PEG 8000 solution (w/v), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **133** (A_{260}, Ɛ = 210500 L•mol^{-1}•cm^{-1}). Separately, **121** (40 pmol, Figure 74) was 5'-phosphorylated on the template (bottom) strand in a 25 μL

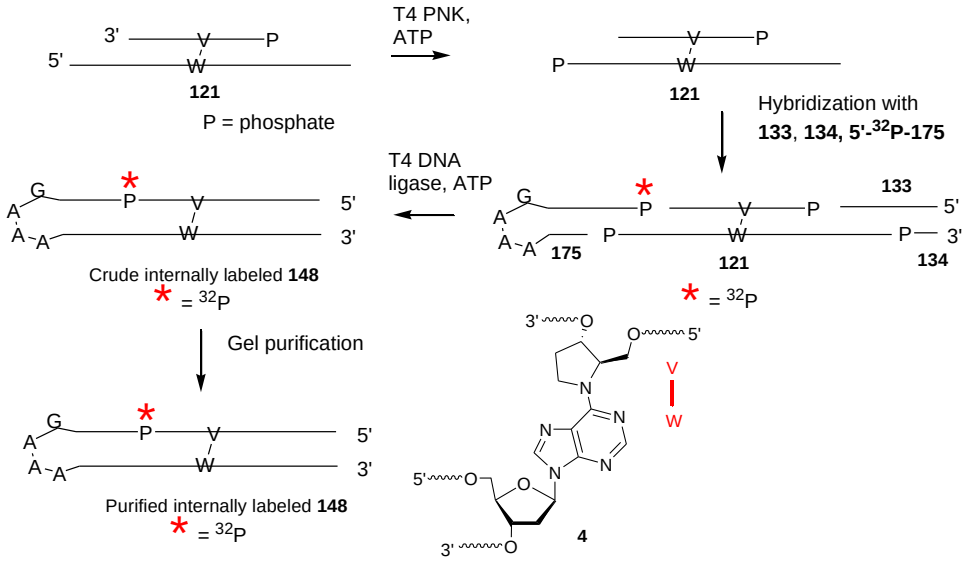
Scheme 46. General Procedure for preparing 5'-^{32}P-t-148



reaction containing 1 × T4 PNK buffer, ATP (80 pmol), and 30 U of T4 PNK at 37 °C for 1 h. Both phosphorylation reactions were stopped by heating at 65 °C for 30 min. The phosphorylated products were hybridized with 160 pmol each of **134** and **175** (Figure 74) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM

EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 5'-^{32}P-t-**148**. The yield (10-20%) of 5'-^{32}P-t-**148** was determined based on the specific activity of 5'-^{32}P-**133**.

Scheme 47. General Procedure for preparing internally labeled 148



General Procedure for Synthesis of Internally ^{32}P-labeled 148. The preparation of internally radiolabeled **148** is described in Scheme 47. Oligonucleotide **175** (80 pmol, Figure 74) was 5'-^{32}P-labeled in reaction (80 µL) containing 1 × T4 PNK exchange buffer buffer (50 mM imidazole-HCl pH 6.4, 18 mM $MgCl_2$, 5 mM DTT, 0.1 mM spermidine, and 0.1 mM ADP), 6% PEG 8000 solution (w/v), 12 µM ATP, 50 µCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was

determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **175** (A_{260}, Ɛ = 366700 L•mol^{-1}•cm^{-1}). Separately, **121** (40 pmol, Figure 74) was 5'-phosphorylated on the template strand in a 25 μL reaction

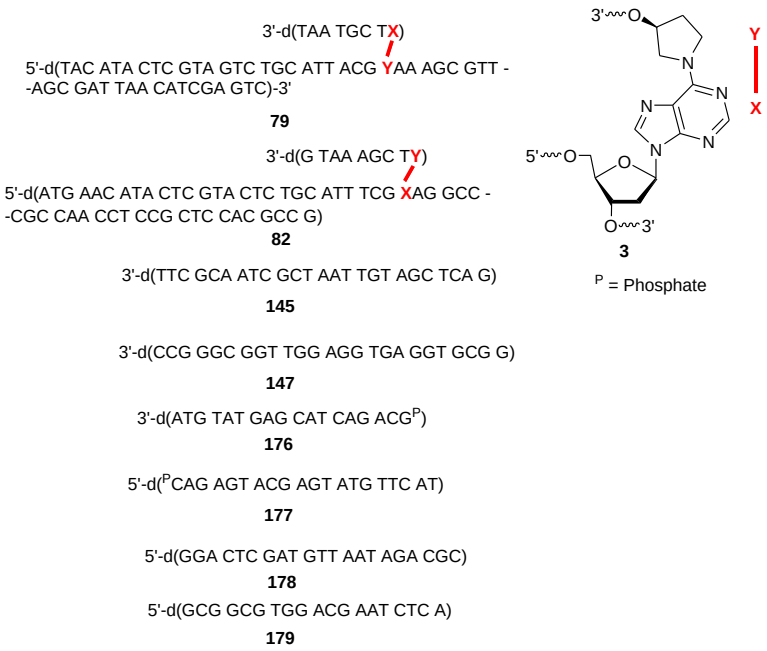


Figure 75. Oligonucleotides used to make **144 and 146**.

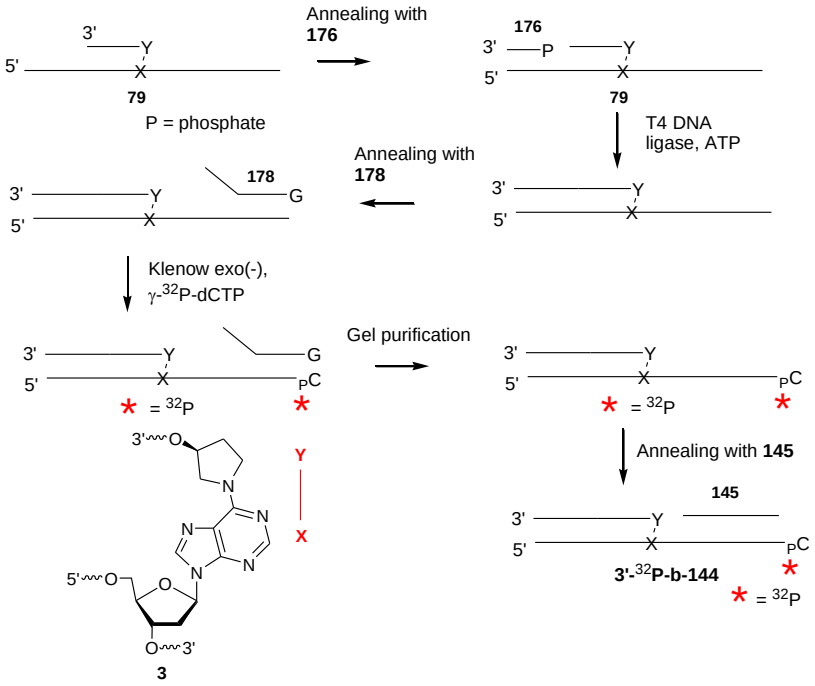
containing 1 × T4 PNK buffer, ATP (80 pmol), and 30 U of T4 PNK at 37 °C for 1 h. Both phosphorylation reactions were stopped by heating at 65 °C for 30 min. The phosphorylated products were hybridized with 160 pmol each of **133** and **134** (Figure 74) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run

under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide internally radiolabeled **148**. The yield (10-20%) of internally ^{32}P-labeled **148** was determined based on the specific activity of 5'-^{32}P-**175**.

General Procedure for Synthesis of 3'-^{32}P-b-144. The preparation of 3'-^{32}P-b-**144** is described in Scheme 48. Oligonucleotide **176** (160 pmol, Figure 75) was hybridized with **79** (50 pmol) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (20 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). The ligation was stopped by heating at 65 °C for 30 min. After addition of **178** (160 pmol, Figure 75) to the solution, it was hybridized with the cross-link by slowly cooling down from 90 °C to 16 °C. The cross-linked substrate was then 3'-^{32}P-labeled in its bottom strand using a "fill in" reaction (150 μL) containing 1 × NEB 2.0 buffer (50 mM NaCl, 10 mM TRIS-hCl, 10 mM MgCl2, 1 mM DTT, pH 7.9), 30 μCi α-^{32}P-dCTP, and 30 U of klenow exo(-) at 37 °C for 3 h. The "fill in" reaction was stopped by heating at 75 °C for 30 min. Excess α-^{32}P-dCTP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and assuming hundred percent recovery of the radiolabeled cross-linked substrate. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the recovered substrate, followed by heating at 90 °C for 5

min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **145** (Figure 75) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 3'-^{32}P-b-**144**. The yield (10-20%) of 3'-^{32}P-b-**144** was determined based on the specific activity of 3'-^{32}P-radiolabeled cross-link, assuming 100% recovery from Sephadex column.

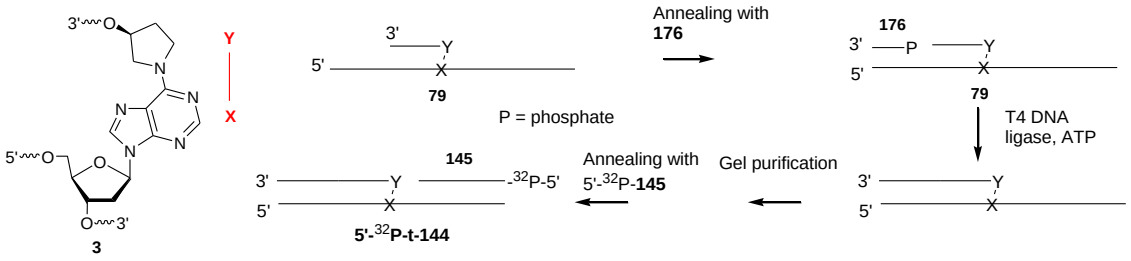
Scheme 48. General Procedure for preparing 3'-^{32}P-b-144



General Procedure for Synthesis of 5'-^{32}P-t-144. The preparation of 5'-^{32}P-t-**144** is described in Scheme 49. Oligonucleotide **79** (1.5 nmol, Figure 75) was hybridized with **176** (3 nmol, Figure 75) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium

chloride solution (50 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.1 cm). The gel was run under limiting power (45 W) for 14 h. The product band was detected using UV, excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was concentrated by repeatedly mixing with 1-butane and then desalted by precipitation. The yield of ligation (22%) was determined from the concentration of isolated product (A_{260}, ε = 633369 L•mol^{-1}•cm^{-1}). In a separate reaction **145** (20 pmol, Figure 75) was 5'-^{32}P-labeled in reaction (80 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. The 5'-^{32}P-**145** was then hybridized with the ligated cross-link by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to obtain 5'-^{32}P-t-**144.**

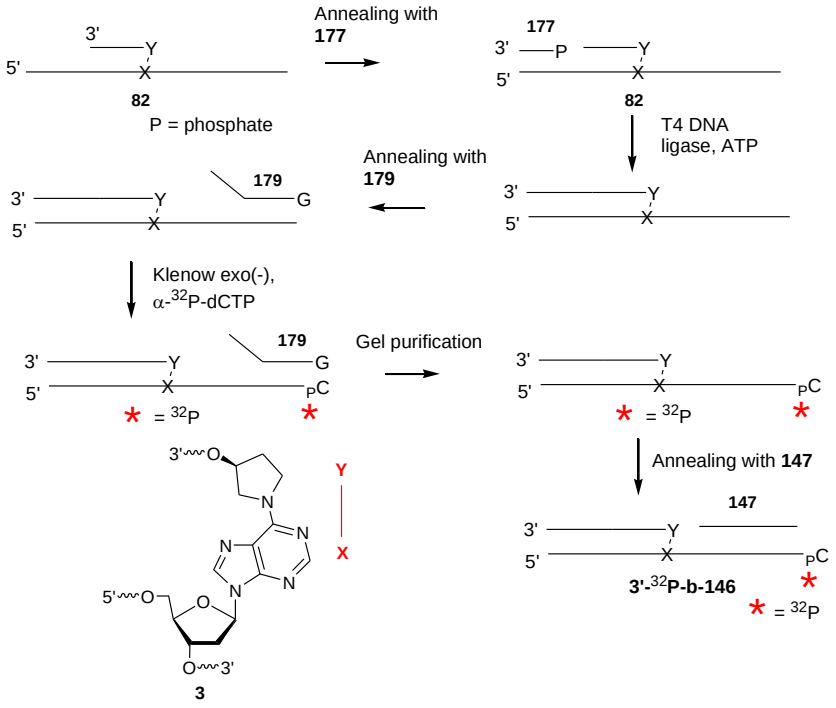
Scheme 49. General Procedure for preparing 5'-^{32}P-t-144



General Procedure for Synthesis of 3'-^{32}P-b-146. The preparation of 3'-^{32}P-b-**146** is described in Scheme 50. Oligonucleotide **177** (160 pmol, Figure 75) was

hybridized with **82** (50 pmol, Figure 75) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (20 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). The

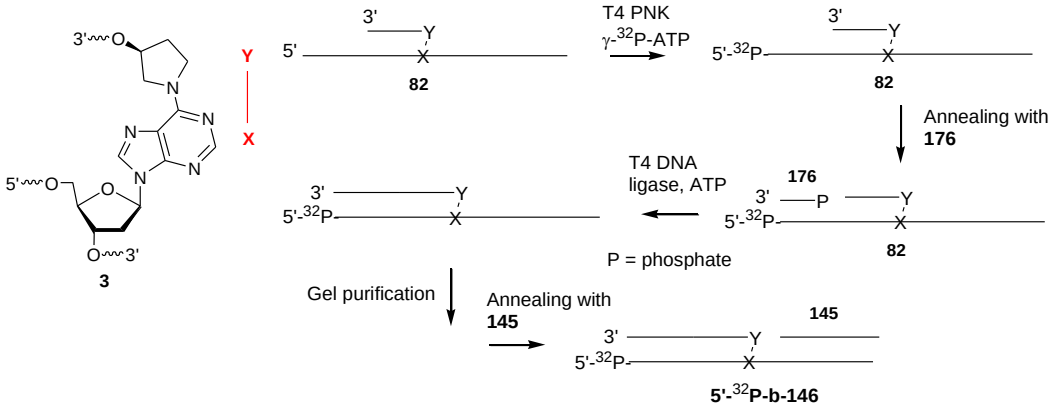
Scheme 50. General Procedure for preparing 3'-^{32}P-b-146



ligation was stopped by heating at 65 °C for 30 min. After addition of **179** (160 pmol, Figure 75) to the solution, it was hybridized with the cross-link by slowly cooling down from 90 °C to 16 °C. The cross-linked substrate was then 3'-^{32}P-labeled in its bottom strand using a "fill in" reaction (150 μL) containing 1 × NEB 2.0 buffer (50 mM NaCl, 10 mM TRIS-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), 30 μCi α-^{32}P-dCTP, and 30 U of Klenow exo(-) at 37 °C for 3 h. The "fill in" reaction was stopped by heating at 75 °C for 30 min. Excess α-^{32}P-dCTP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the

radioactivity (using a liquid scintillation counter) and assuming hundred percent recovery of the radiolabeled cross-linked substrate. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the recovered substrate, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **147** (Figure 75) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 3'-^{32}P-b-**146**. The yield (10-20%) of 3'-^{32}P-b-**146** was determined based on the specific activity of 3'-^{32}P-radiolabeled cross-link, assuming 100% recovery from the Sephadex column.

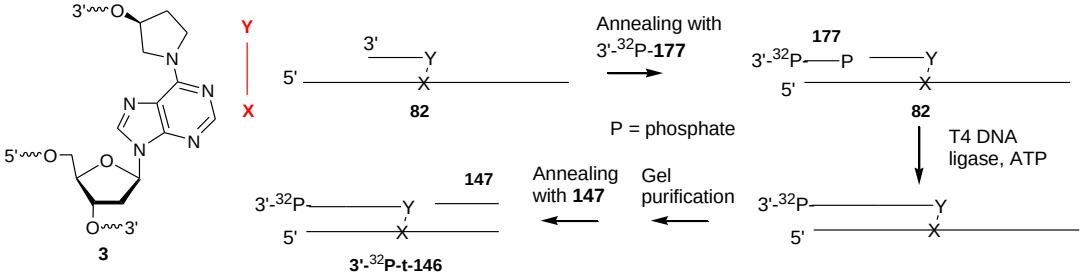
Scheme 51. General Procedure for preparing 5'-^{32}P-b-146



General Procedure for Synthesis of 5'-^{32}P-b-146. The preparation of 5'-^{32}P-b-**9** is described in Scheme 51. Oligonucleotide **82** (50 pmol, Figure 75) was 5'-^{32}P-labeled in reaction (80 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. The

phosphorylation reaction was stopped by heating at 65 °C for 30 min. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **82** (A_{260}, Ɛ = 633929 L•mol^{-1}•cm^{-1}). The ^{32}P-radiolabeled **82** were hybridized with **176 (**160 pmol, Figure 75) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **145** (Figure 75) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 5´-^{32}P-b-**146**. The yield (10-20%) of 5´-^{32}P-b-**146** was determined based on the specific activity of 5´-^{32}P-**82**.

Scheme 52. General Procedure for preparing 3'-^{32}P-t-146



General Procedure for Synthesis of 3'-^{32}P-t-146. The preparation of 3'-^{32}P-t-**146** is described in Scheme 52. Oligonucleotide **177** (80 pmol, Figure 75) was 3'-^{32}P-labeled in reaction (80 μL) containing 1 × TdT buffer buffer (50 mM KOAc, 20 mM TRIS-OAc, 10 mM $MgCl_2$, pH 7.9), 40 μCi α-^{32}P-cordycepin triphosphate, and 30 U of terminal deoxynucleotidyl transferase at 37 °C for 2 h. The reaction was stopped by heating at 75 °C for 30 min. Excess α-^{32}P-cordycepin triphosphate was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **177** (A_{260}). The 3'-^{32}P-labeled **177** was hybridized with **82** (50 pmole, Figure 75) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (100 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 18 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **147** (Figure 75) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 3'-^{32}P-t-**146**. The yield (10-20%) of 3'-^{32}P-t-**146** was determined based on the specific activity of 3'-^{32}P-**177**.

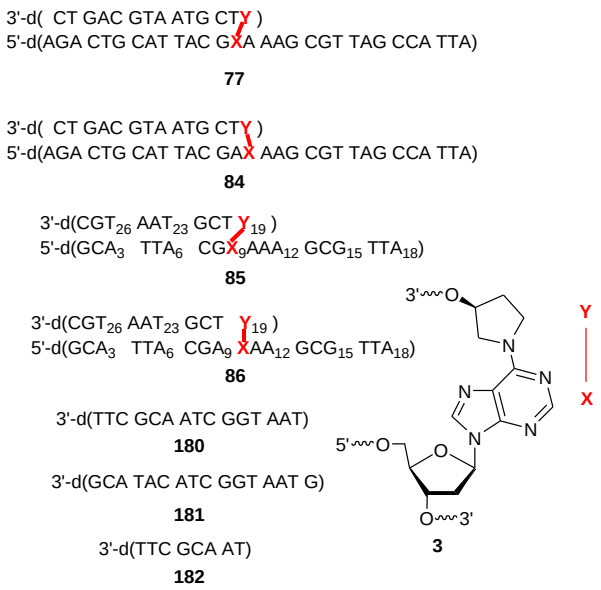
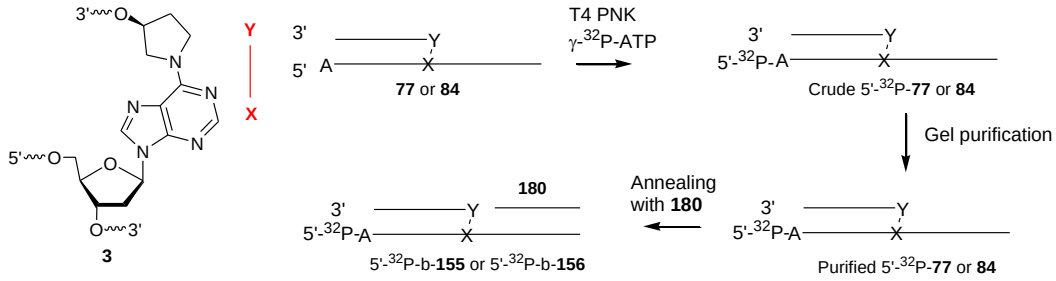


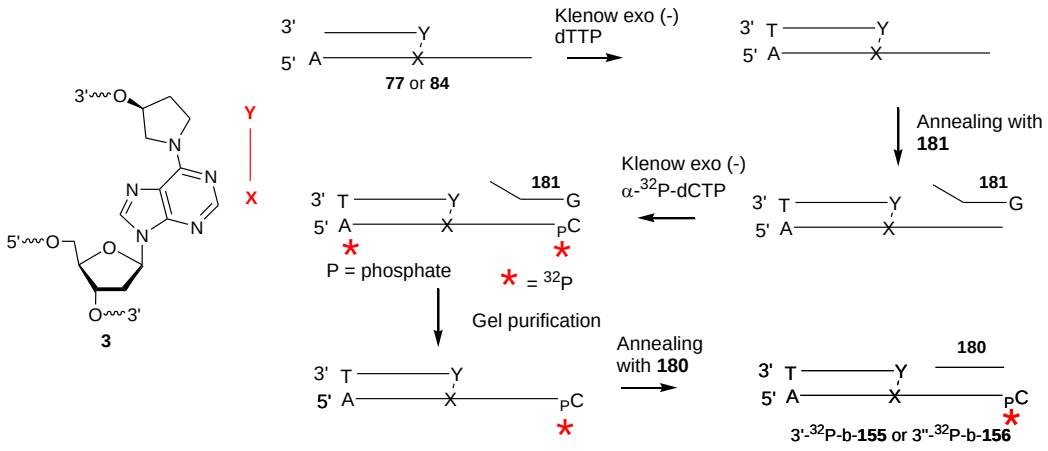
Figure 76. Oligonucleotides used to make **155**, **156, 160** and **161.**

Scheme 53. General Procedure for preparing 5'-^{32}P-b-155 and 5'-^{32}P-b-156



General Procedure for Synthesis of 5'-^{32}P-b-155 and 5'-^{32}P-b-156. Scheme 53 describes the preparation of 5'-^{32}P-b-**155** and 5'-^{32}P-b-**156** from **77** and **84**, respectively. Oligonucleotide **77** or **84** (50 pmol, Figure 76) was 5'-^{32}P-labeled in reaction (80 µL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 50 µCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. The phosphorylation reaction was stopped by heating at 65 °C for 30 min. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL

Scheme 54. General Procedure for preparing 3'-^{32}P-b-155 and 3'-^{32}P-b-156

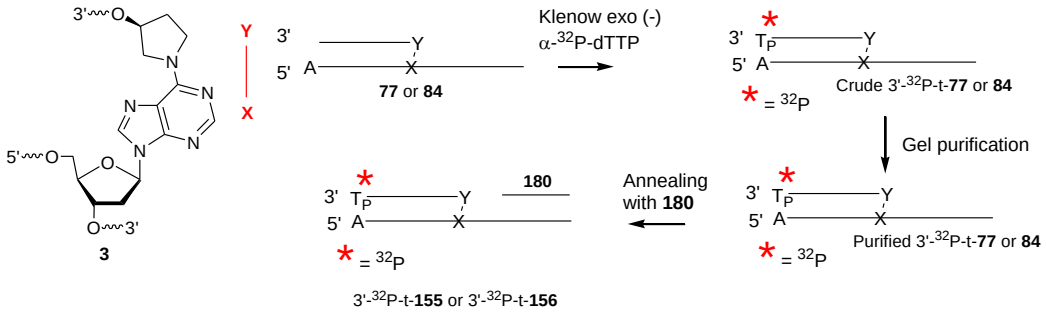


Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and assuming hundred percent recovery of the radiolabeled cross-linked substrate. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the recovered substrate, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 8 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **180** (Figure 76) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 5'-^{32}P-b-**155** and 5'-^{32}P-b-**156**. The yield (30-50%) of 3'-^{32}P-b-**146** was determined based on the specific activity of 5'-^{32}P-labeled **77** and **84**.

General Procedure for Synthesis of 3'-^{32}P-b-155 and 3'-^{32}P-b-156. Scheme 54 describes the preparation of 3'-^{32}P-b-**155** and 3'-^{32}P-b-**156** from **77** and **84**, respectively. Oligonucleotide **77** or **84** (50 pmol, Figure 76) was subjected to a "fill in" reaction (30 μL) containing 1 × NEB 2.0 buffer (50 mM NaCl, 10 mM TRIS-hCl, 10 mM MgCl2, 1

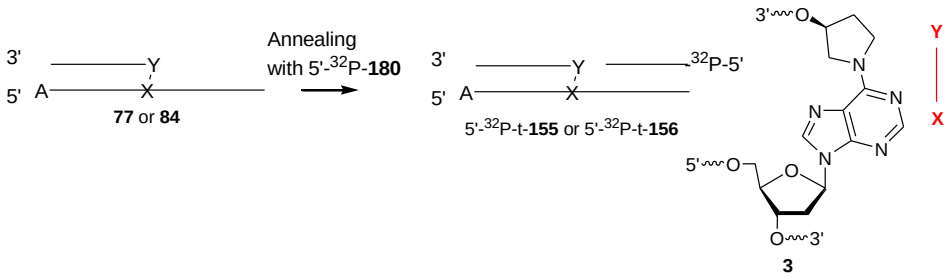
mM DTT, pH 7.9), dTTP (1 mM), and 30 U of Klenow exo(-) at 37 °C for 3 h. The "fill in" reaction was stopped by heating at 75 °C for 30 min. The reaction was then added and hybridized with **181** (100 pmol, Figure 76) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (100 μL) by slowly cooling from 90 °C to 16 °C. The annealed substrate was then 3'-^{32}P-labeled in its bottom strand using a "fill in" reaction (150 μL) containing 1 × NEB 2.0 buffer (50 mM NaCl, 10 mM TRIS-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), 30 μCi α-^{32}P-dCTP, and 30 U of Klenow exo(-) at 37 °C for 3 h. The "fill in" reaction was stopped by heating at 75 °C for 30 min. Excess α-^{32}P-dCTP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and assuming hundred percent recovery of the radiolabeled cross-linked substrate. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the recovered substrate, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 8 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **180** (Figure 76) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 3'-^{32}P-b-**155** and 3'-^{32}P-b-**156**. The yield (30-50%) of 3'-^{32}P-b-**146** was determined based on the specific activity of 3'-^{32}P-labeled **77** and **84.**

Scheme 55. General Procedure for preparing 3'-^{32}P-t-155 and 3'-^{32}P-t-156



General Procedure for Synthesis of 3'-^{32}P-t-155 and 3'-^{32}P-t-156. Scheme 55 describes the preparation of 3'-^{32}P-t-**155** and 3'-^{32}P-t-**156** from **77** and **84**, respectively. Oligonucleotide **77** or **84** (50 pmol, Figure 76) was subjected to a "fill in" reaction (50 μL) containing 1 × NEB 2.0 buffer (50 mM NaCl, 10 mM TRIS-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), 30 μCi α-^{32}P-dTTP, and 30 U of Klenow exo(-) at 37 °C for 3 h. The "fill in" reaction was stopped by heating at 75 °C for 30 min. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and assuming hundred percent recovery of the radiolabeled cross-linked substrate. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the recovered substrate, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 8 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge, concentrated and then hybridized with **180** (Figure 76) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 3'-^{32}P-t-**155** and 3'-^{32}P-t-**156**. The yield (30-50%) of 3'-^{32}P-b-**146** was determined based on the specific activity of 3'-^{32}P-labeled **77** and **84.**

Scheme 56. General Procedure for preparing 5'-^{32}P-t-155 and 5'-^{32}P-t-156



General Procedure for Synthesis of 5'-^{32}P-t-155 and 5'-^{32}P-t-156. Scheme 56 describes the preparation of 5'-^{32}P-t-**155** and 5'-^{32}P-t-**156** from **77** and **84** (Figure 76), respectively. Oligonucleotide **180** (30 pmol, Figure 76) was 5'-^{32}P-labeled in reaction (50 µL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 50 µCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. The phosphorylation reaction was stopped by heating at 65 °C for 30 min. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the absorption (A_{260}) of 5'-^{32}P-**180**. Oligonucleotide 5'-^{32}P-**180** was hybridized with with **77** or **84** (60 pmol, Figure 76) by slow cooling from 90 °C to 25 °C in 1 × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) to provide 5'-^{32}P-t-**155** and 5'-^{32}P-t-**156**.

General Procedure for Preparing 160 and 161 for T_m Experiments. Oligonucleotides **85** and **86** (500 pmol, Figure 76) were independently mixed with **182** (500 pmol, Figure 76) in a solution (500 µL) containing 10 mM PIPES (pH 7.0), NaCl (100 mM) and $MgCl_2$ (10 mM). The oligonucleotide solutions were hybridized by slow cooling from 90 °C to 25 °C and then dispensed into three UV quartz cuvettes (160 µL each).

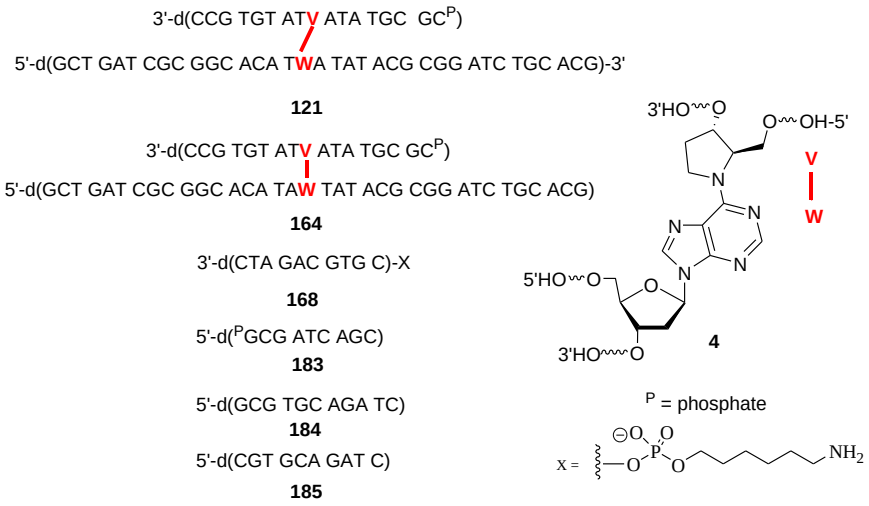
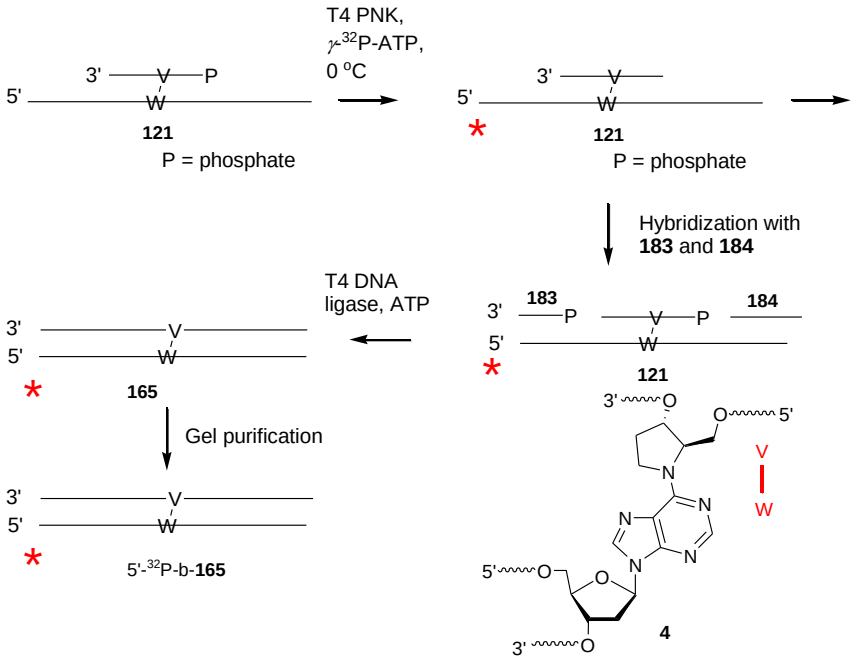


Figure 77. Oligonucleotides used to make **165**, **166** and **169**.

General Procedure for Preparing 5'-^{32}P-b-165. Scheme 57 describes the general procedure for preparing 5'-^{32}P-b-**165** from **121** (Figure 77). Oligonucleotide **121** (80 pmol, Figure 77) was 5'-^{32}P-labeled in reaction (50 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 6% PEG 8000 solution (w/v), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 0 °C for 2 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The phosphorylation reaction was stopped by heating at 65 °C for 30 min and the T4 PNK was removed by phenol-chloroform-isopropanol extraction. The 5'-^{32}P-b-**121** was hybridized with **183** and **183** (160 pmol each, Figure 77) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (120 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the reaction, followed by heating at 90 °C for 5 min and cooling on ice prior to purification

by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 11 h.

Scheme 57. General procedure for preparing 5'-^{32}P-b-165

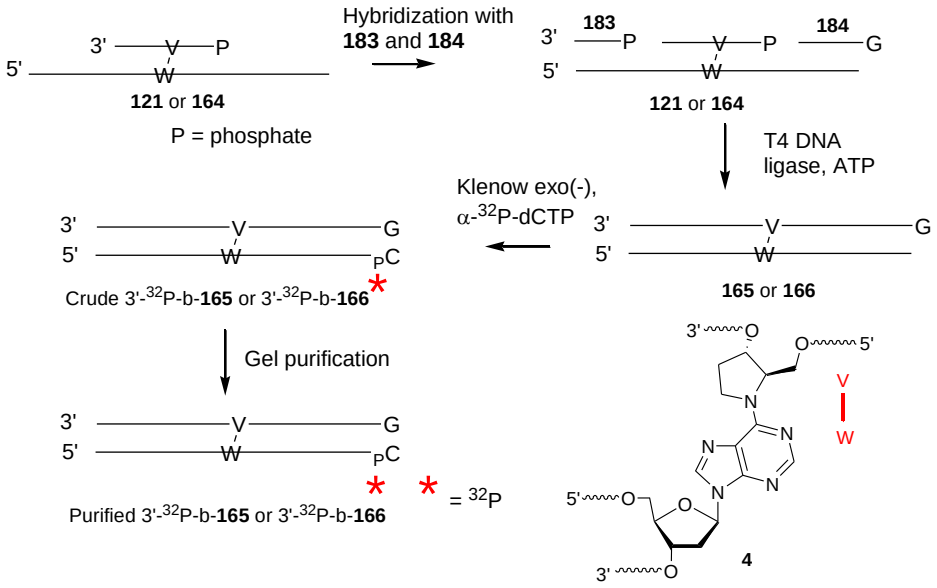


The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 5'-^{32}P-b-**165**. The isolated final products were suspended in × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) and hybridized by slow cooling from 90 °C to 25 °C.

General Procedure for Preparing 3'-^{32}P-b-165 and 3'-^{32}P-b-166. Scheme 58 describes the general procedure for preparing 3'-^{32}P-b-**165** and 3'-^{32}P-b-**166** from **121** and **164** (Figure 77), respectively. Oligonucleotide **121** or **164** (50 pmol) was annealed with **183** and **184** (100 pmol, Figure 77) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (50 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase

buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). The ligated substrates were subjected to a "fill in" reaction (100 μL) containing 1 × NEB 2.0 buffer (50 mM NaCl, 10 mM TRIS-HCl, 10 mM $MgCl_2$, 1 mM DTT, pH 7.9), dTTP (1 mM), and 30 U of Klenow exo(-) at 37 °C for 3 h. The "fill in" reaction

Scheme 58. General procedure for preparing 3'-^{32}P-b-165 and 3'-^{32}P-b-166



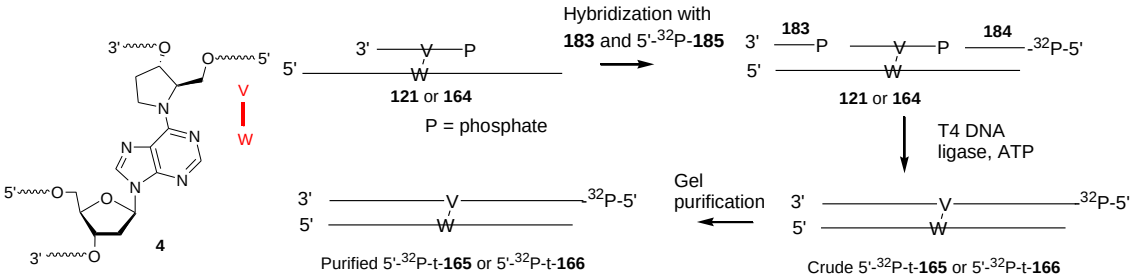
was stopped by heating at 75 °C for 30 min. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and assuming hundred percent recovery from Sephadex column. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added the ^{32}P-labeled substrates, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 11 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 5 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 3'-^{32}P-b-**165** and 3'-^{32}P-b-**166**. The isolated

final products were suspended in × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) and hybridized by slow cooling from 90 °C to 25 °C. The yield (4-10%) of purified 3'-^{32}P-b-**165** and 3'-^{32}P-b-**166** was determined based on the specific activity of crude substrates.

General Procedure for Preparing 5'-^{32}P-t-165 and 5'-^{32}P-t-166. Scheme 59 describes the general procedure for preparing 5'-^{32}P-t-**165** and 5'-^{32}P-t-**166** from **121** and **164** (Figure 77), respectively. Oligonucleotide **185** (80 pmol, Figure 77) was 5'-^{32}P-labeled in reaction (50 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 6% PEG 8000 solution (w/v), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **185** (A_{260}, ε = 94900 L•mol^{-1}•cm^{-1}). The phosphorylation reaction was stopped by heating at 65 °C for 30 min. The 5'-^{32}P-**185** was hybridized with 50 pmol of **121** or **164 (**Figure 77) and 160 pmol of **183** (Figure 77) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (120 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the reaction, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 11 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The

supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 5'-^{32}P-t-**165** and 5'-^{32}P-t-**166**. The isolated final products were suspended in × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) and hybridized by slow cooling from 90 °C to 25 °C. The yield (4–10%) of 5'-^{32}P-t-**165** and 5'-^{32}P-t-**166** was determined based on the specific activity of 5'-^{32}P-**185**.

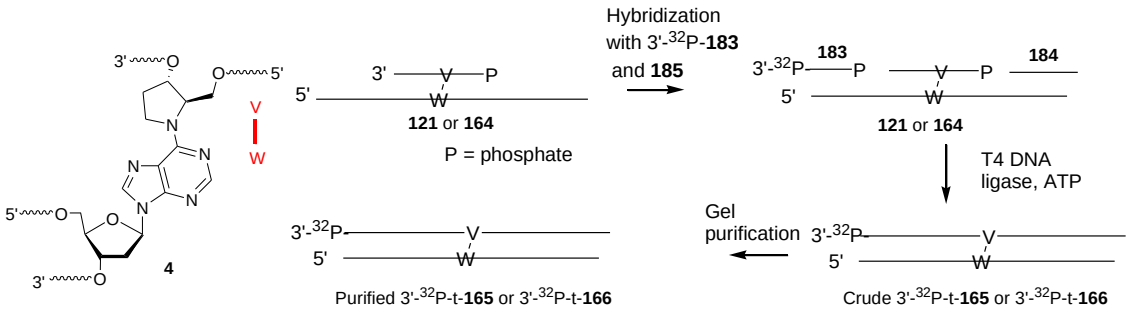
Scheme 59. General procedure for preparing 5'-^{32}P-t-165 and 5'-^{32}P-t-166



General Procedure for Preparing 3'-^{32}P-t-165 and 3'-^{32}P-t-166. Scheme 60 describes the general procedure for preparing 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166** from **121** and **164** (Figure 77), respectively. Oligonucleotide **183** (80 pmol, Figure 77) was 3'-^{32}P-labeled in reaction (50 µL) containing containing 1 × TdT buffer buffer (50 mM KOAc, 20 mM TRIS-OAc, 10 mM $MgCl_2$, pH 7.9), 40 µCi α-^{32}P-cordycepin triphosphate, and 30 U of terminal deoxynucleotidyl transferase at 37 °C for 2 h. Excess α-^{32}P-cordycepin triphosphate was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of **183** (A_{260}, Ɛ = 86300 L•mol^{-1}•cm^{-1}). The ^{32}P-labeled **22** was hybridized with **121** or **164** (50 pmole) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (100 µL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10

mM DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the reaction, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 11 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166**. The isolated final products were suspended in × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) and hybridized by slow cooling from 90 °C to 25 °C. The yield (4–10%) of 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166** was determined based on the specific activity of 3'-^{32}P-**183**.

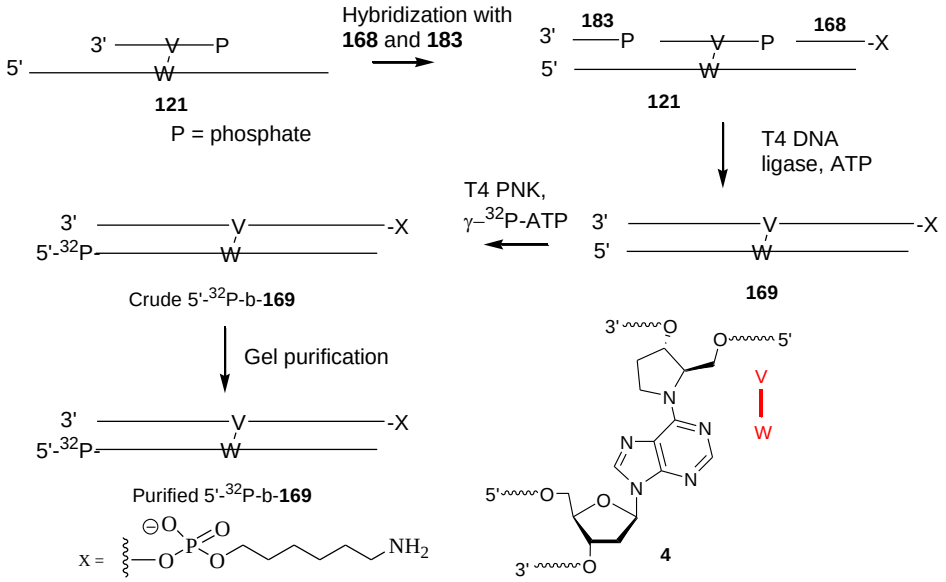
Scheme 60. General procedure for preparing 3'-^{32}P-t-165 and 3'-^{32}P-t-166



General Procedure for Synthesis of 5'-^{32}P-b-169. Scheme 61 describes the general procedure for preparing 3'-^{32}P-b-**169** from **121** (Figure 77). Oligonucleotide **121** (50 pmol) was annealed with **168** and **183** (100 pmol, Figure 77) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (50 μL) by slowly cooling from 90 °C to 16 °C. The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM DTT, 1 mM ATP). The ligated substrates were subjected to polynucleotide kinase

phosphorylation reaction (50 μL) containing 1 × T4 PNK buffer (70 mM TRIS-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT), 6% PEG 8000 solution (w/v), 50 μCi γ-^{32}P-ATP, and 30 U of T4 PNK at 37 °C for 2 h. Excess γ-^{32}P-ATP was removed by passing the reaction through a 1 mL Sephadex G25 column. The phosphorylation reaction was

Scheme 61. General procedure for preparing 5'-^{32}P-b-169



stopped by heating at 65 °C for 30 min. Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added the ^{32}P-labeled substrate, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 11 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 5 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge and concentrated to provide 5'-^{32}P-b-**169.** The isolated final product was suspended in × PBS buffer (10 mM sodium phosphate, pH 7.2, 100 mM NaCl) and hybridized by slow cooling from 90 °C to 25 °C.

General Procedure for Preparing 129 and 130 for T_m Experiments. Oligonucleotides **129** and **130** (500 pmol, Figure 77) were independently suspended in solutions (500 μL) containing 10 mM PIPES (pH 7.0), NaCl (100 mM) and $MgCl_2$ (10 mM). The oligonucleotide solutions were hybridized by slow cooling from 90 °C to 25 °C and then dispensed into three UV quartz cuvettes (160 μL each).

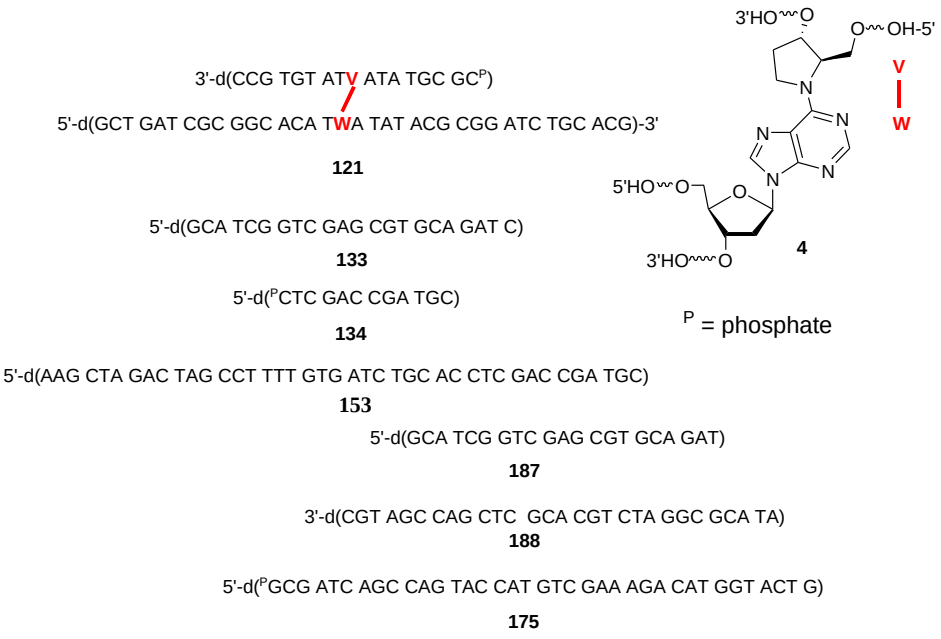
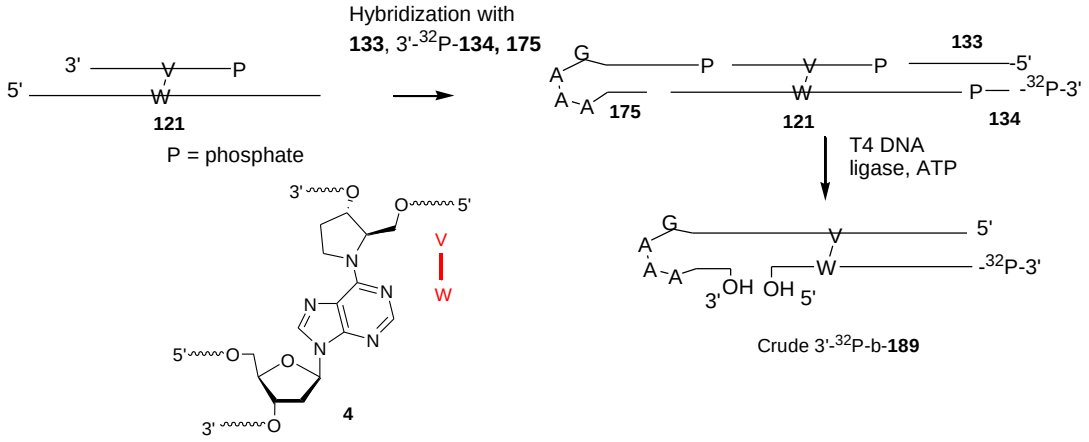


Figure 78. Oligonucleotides used for making 5'-^{32}P-t-**151**, 5'-^{32}P-t-**152**, and 3'-^{32}P-b-**189**

General Procedure for Synthesis of 3'-^{32}P-b-189. Scheme 62 describes the general procedure for preparing 3'-^{32}P-b-**189** from **121** (Figure 78). Oligonucleotide **134** (80 pmol, Figure 78) was 3'-^{32}P-labeled in reaction (80 μL) containing containing 1 × TdT buffer buffer (50 mM KOAc, 20 mM TRIS-OAc, 10 mM $MgCl_2$, pH 7.9), 40 μCi α-^{32}P-cordycepin triphosphate, and 30 U of terminal deoxynucleotidyl transferase at 37 °C for 2 h. Excess α-^{32}P-cordycepin triphosphate was removed by passing the reaction through a 1 mL Sephadex G25 column. The specific activity was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration of

134 (A_{260}, ε = 108100 L•mol^{-1}•cm^{-1}). The ^{32}P-labeled **134** was hybridized with **121** (50 pmole), **133** (160 pmol), and **153** (160 pmol) in 10 mM sodium phosphate (pH 7.2) and 100 mM sodium chloride solution (100 μL) by slowly cooling from 90 °C to 16 °C (Figure 78). The reaction was incubated overnight at 16 °C in presence of T4 DNA ligase (50 U) and 1 × ligase buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM

Scheme 62. General procedure for making 3'-^{32}P-b-189



DTT, 1 mM ATP). Formamide loading buffer (50 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to the reaction, followed by heating at 90 °C for 5 min and cooling on ice prior to purification by 20% denaturing PAGE (37 × 32 × 0.04 cm). The gel was run under limiting power (55 W) for 11 h. The product band was excised from the gel, crushed, and the DNA was eluted overnight in 1 mL of elution buffer (0.2 M NaCl and 1 mM EDTA). The supernatant was desalted using a 100 mg C-18-Sep-Pak cartridge to provide 3'-^{32}P-b-**189**. The yield (4–10%) of 3'-^{32}P-b-**189** was determined based on the specific activity of 3'-^{32}P-**134**.

General Procedure for Restriction Enzyme Treatment of 135. The restriction enzyme sites on **135** are described in App. Figure 17 and 55. The appropriate restriction

enzyme (10 U) and 5´-^{32}P-b-**135** were incubated in a 10 μL reaction containing of 1 X CutSmart buffer (50 mM potassium acetate, 20 mM TRIS acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) at 37°C (RsaI) or 65°C (Taq$^{\alpha}$I) for 1 h, at which time formamide loading buffer (2 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added followed by heating at 90°C and cooling on ice. The reactions were analyzed by 20% denaturing PAGE.

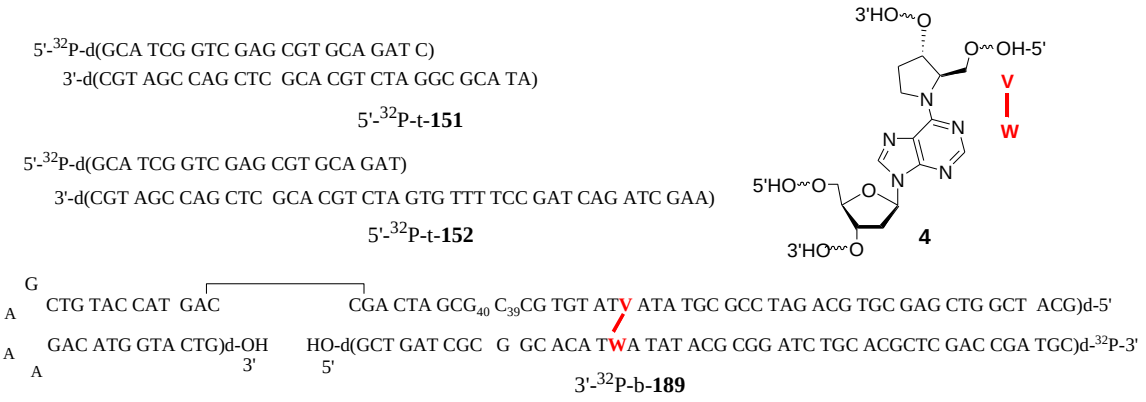


Figure 79. Sequences of 5'-^{32}P-t-**151**, 5'-^{32}P-t-**152**, and 3'-^{32}P-b-**189.**

General Procedure for Restriction Enzyme Treatment of 170. The restriction enzyme sites on **170** are described on App. Figure 55. The appropriate restriction enzyme (10 U) and 5'-^{32}P-b-**170** were incubated in a 10 μL reaction containing of 1 X CutSmart buffer (50 mM potassium acetate, 20 mM TRIS acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) at 37°C (RsaI) or 65°C (Taq$^{\alpha}$I) for 1 h, at which time formamide loading buffer (2 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added followed by heating at 90°C and cooling on ice. The reactions were analyzed by 15% denaturing PAGE.

General Procedure of Restriction Enzyme Treatment for Characterization of 146. The appropriate restriction enzyme (10 U) and ^{32}P-labeled **146** were incubated in a 8

μL reaction containing of 1 × CutSmart buffer (50 mM potassium acetate, 20 mM TRIS acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) at 37°C (Fnu4HI and BsrBI) for 1 h, at which time formamide loading buffer (2 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added followed by heating at 90°C and cooling on ice. The reactions were analyzed by 15% denaturing PAGE.

General Procedure of Restriction Enzyme Treatment for Characterization of 5'-^{32}P-t-148. The restriction enzyme Taq$^{\alpha}$I (10 U) and 5´-^{32}P-t-**148** were incubated in a 8 μL reaction containing of 1 × CutSmart buffer (50 mM potassium acetate, 20 mM TRIS acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) at 65°C (Taq$^{\alpha}$I) for 1 h. In two independent experiments, CviQI was incubated with 5´-^{32}P-t-**8** or 5´-^{32}P-t-**11** in 8 μL reactions containing of 1 × NEB 3.1 buffer (50 mM potassium acetate, 20 mM TRIS acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) at room temperature for 1 h. After 1 h, formamide loading buffer (2 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added to these reactions followed by heating at 90°C and cooling on ice. The reactions were analyzed by 15% denaturing PAGE.

General Procedure of Restriction Enzyme Treatment for Characterization of Internally Radiolabeled 148. The appropriate restriction enzyme (10 U) and internally radiolabeled **148** were incubated in a 8 μL reaction containing of 1 × CutSmart buffer (50 mM potassium acetate, 20 mM TRIS acetate, 10 mM magnesium acetate, 100 μg/mL BSA, pH 7.9) at 37°C (Fnu4HI and HpyCH4V) for 1 h, at which time formamide loading buffer (2 μL, 90% formamide, 10 mM EDTA, pH 8.0) was added followed by heating at 90°C and cooling on ice. The reactions were analyzed by 15% denaturing PAGE.

General Procedure of Determination of Thermal Stability of 5'-^{32}P-t-144 Under Nucleotide Excision Repair Conditions. Oligonucleotide 5'-^{32}P-t-**144** (2 nM) was incubated with **145** (10 nM) for 1 h in 60 μL reaction containing 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 3 h. Aliquots (9 μL) were taken out at 0, 5, 10, 15, 30 and 60 min and mixed with glycerol loading buffer (2 μL, 80% glycerol in water, v/v). The exchange was analyzed by 20% non-denaturing PAGE.

General Procedure of Determination of Thermal Stability of 5'-^{32}P-t-146 Under Nucleotide Excision Repair Conditions. Oligonucleotide 5'-^{32}P-t-**146** (2 nM) was incubated with **147** (10 nM) for 1 h in 60 μL reaction containing 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 3 h **.** Aliquots (9 μL) were taken out at 0, 5, 10, 15, 30 and 60 min and mixed with glycerol loading buffer (2 μL, 80% glycerol in water, v/v). The exchange was analyzed by 20% non-denaturing PAGE.

General procedure of determination of thermal stability of 151 under nucleotide excision repair condition. Oligonucleotide 5'-^{32}P-t-**151** (0.26 nM) was incubated with **153** (0.26 nM or 2.6 nM) in a 50 μL reaction containing 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM DTT, and ATP 1 mM) at 55 °C for 3 h. The experiment was cooled down to 4 °C and added with glycerol loading buffer (5.5 μL, 80% glycerol in water, v/v). The exchange was analyzed by 15% non-denaturing PAGE.

UvrABC Incision Reaction. The purified cross-linked DNA was resuspended in 100 mM NaCl and 10 mM potassium phosphate buffer (pH 7.2, 25 μL), and rehybridized by heating to 65 °C (2 min), cooled to room temperature over the course of 2 h, and allowed to equilibrate at 4 °C overnight. The UvrA, UvrB, and UvrC were freshly prepared from stock solutions and heated individually at 65 °C for 10 min before use. They were added sequentially to the reactions. The reaction buffer contains 50 mM Tris-HCl (pH 7.5), $MgCl_2$ (10 mM), KCl (50 mM), DTT (5 mM), and ATP (1.0 mM). The cross-linked DNA (2.0 nM) was incubated with UvrA (20 nM), UvrB (100 nM), and UvrC (50 nM) in a total volume of 20 μL at 55 °C. The concentration of cross-linked DNA was based upon the specific activity of the initially labeled oligonucleotide, which was determined by counting the radioactivity (using a liquid scintillation counter) and measuring the concentration (A_{260}). After 60 min the reaction was quenched by precipitation with 5 M NH_4OAc (5 μL), 1 μg/μL calf thymus DNA (5 μL), and cold ethanol (75 μL). The incision products were separated by 20% denaturing PAGE and visualized using the Phosphorimager. For the time course reaction (60 μL total volume) aliquots (7 μL) were removed at the prescribed times and immediately quenched by addition of 90% formamide loading buffer (3 μL) and heating at 90°C for 2 min.

Gel Shift Assays. Gel shift assays were performed to determine the Kd values of UvrA for ICL-containing DNA.[125] Reaction contained 0.3 nM DNA duplex (**144**, **146** and **149**) ^{32}P-labeled at the end in 1 × NER buffer [50 mM Tris-HCl (pH 7.5), $MgCl_2$ (10 mM), KCl (50 mM), DTT (5 mM), and ATP (1.0 mM)], and UvrA concentration ranging from 0.5 nM to 200 nM. Samples of protein/DNA mixture was incubated at 55 °C for 20

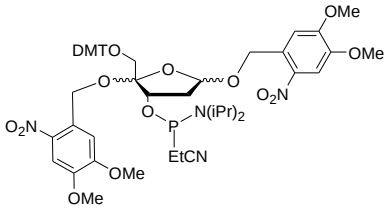
min followed by the addition of 2 μL of 80% glycerol. Bound versus unbound DNA was visualized using electrophoresis on a 4% nondenaturing PAGE (80:1 acrylamide:bisacrylamide ratio) at 4 °C in TBE running buffer (89 mM Tris, 89 mM boric acid, 2 mM ETDA) containing 10 mM $MgCl_2$ and 1 mM ATP at 150 V for 2 – 3 h. Gels were dried and exposed to a storage phosphor screen overnight. K_d values were determined by fitting the data (percent bound substrate vs. log[UvrA]) in one site binding Hill equation ($\theta = \frac{[L]^n}{K_d + [L]^n}$, where θ is fraction of DNA bound to protein, [L] is free protein concentration, n is the number of binding sites, K_d is dissociation constant) in Origin 8.0. The K_d values were determined from data generated from at least three separate experiments using separately diluted UvrA in each experiment.

The loading of UvrB by UvrA was examined by incubating ^{32}P-labeled 0.3 nM DNA duplex (**144**, **146** and **149**) with UvrA (0.5,) in the presence of 500 nM UvrB in 1 × NER buffer [50 mM Tris-HCl (pH 7.5), $MgCl_2$ (10 mM), KCl (50 mM), DTT (5 mM), and ATP (1.0 mM)] at 55 °C for 20 min. After addition of 80% glycerol, samples were electrophoresed as described above.[125] The percent UvrB:DNA complex was determined from quantitation of the storage phosphor autoradiogram. At least three separate experiments were performed and the values averaged to provide the % UvrB loaded onto the damaged DNA.

UV Melting Temperature Experiments. UV melting temperature experiments were carried out in quartz cells with a path length of 1 cm. The samples were prepared by mixing oligonucleotides (1 μM), PIPES, pH 7.0 (10 mM), NaCl (100 mM), and $MgCl_2$ (10 mM). For substrates During the forward melting experiments on **85**, **86**, **160** and **161**,

the temperature was held at 13 °C to 2 min, then ramped linearly to 85 °C at a rate 1.0 °C / min while monitoring the change in UV absorbance at 260 nm. During the reverse experiment, the temperature was held at 85 °C to 5 min, then ramped linearly to 13 °C at a rate 1.0 °C / min while monitoring the change in UV absorbance at 260 nm. In addition, the cuvette rack was constantly flushed with N_2 flow to avoid moisture condensation at low (< 17 °C). For substrates **129** and **130**, the low and high temperature during forward and reverse experiments were kept at 20 °C and 85 °C, respectively. For each experiment, the change in absorbance (260 nm) with temperature for a solution containing the same buffer but without the oligonucleotides was monitored and subsequently subtracted from the absorbance (260 nm) of the oligonucleotides. The data was processed using Origin 6.1 by plotting the absorbance (260 nm) versus temperature and fitting the curve to the Boltzmann equation [y = A2 + (A1-A2)/(1 + exp((x-x0)/dx)), where y is the absorbance, A1 is the intial value (left horizontal asymptote), A2 is the final value (right horizontal asymptote), x0 is center (point of inflection or T_m), dx is the change in X corresponding to the most significant change in Y value].

Synthesis of 162 Containing C4-AP Precursor. Synthesis, ICL formation and determination of ICL location was done following procedures published earlier.[8] Standard synthesis cycles (25 s coupling, 5 s capping with acetic anhydride, 15 s oxidation with 1 M t-butylhydroperoxide in THF, 95 s detritylation with 3% TCA in methylene chloride) were employed prior to the incorporation of the photochemical precursor of the lesion. After the C4-AP phosphoramidite was inserted, the cycles were modified to include a shorter detritylation period (45 s) to



C4-AP phosphoramidite

ensure stability of the lesion. The C4-AP phosphoramidite was itself incorporated with 5 min coupling, 25 s capping, 40 s oxidation, and the shorter 45 s detritylation using the same synthesis reagents. After overnight aqueous ammonia deprotection (55°C), **162** were purified by 20% PAGE.

General Procedure for ICL formation in C4-AP Containing Duplexes. Duplex DNA (**163**, 200 nM) 3'-^{32}P-labelled at the complementary strand in PBS buffer (100 mM NaCl and 10 mM sodium phosphate, pH 7.2) was photolyzed in a clear eppendorf for 45 min at room temperature. The DNA was then diluted to 100 nM. These reactions were typically incubated at 37 °C for 10 h or overnight. At the end of incubation period, the reaction solution was mixed with equal volume of formamide loading buffer and the DNA was resolved using 20% denaturing polyacrylamide gel electrophoresis (PAGE).

Determination of ICL Location using the Hydroxyl-Radical Cleavage Reaction. Duplex DNA (**163**, concentration 100 nM) constructed of a C4-AP containing strand and a 3′-^{32}P-labeled complement was incubated at 37°C overnight in PBS. The reaction was evaporated and purified by 20% analytical (0.4 mm thickness) PAGE. The bands corresponding to the high molecular weight ICL product was carefully cut from the gel, crushed, and eluted in 500 µL of water for 1 h at room temperature with mild vortexing. The solution was then filtered using a 10 mL Polyprep Column (Biorad) and desalted using a C18 column (100 mg Sep-Pak, Waters). After concentration, a portion of the desalted DNA (30,000- 60,000 cpm) was resuspended in 8 µL of water and mixed with 10 µL of the 2 × oxidation buffer (20 mM NaCl, 20 mM sodium phosphate, pH 7.2, 2 mM sodium ascorbate, and 0.25 mM hydrogen peroxide). To the reaction 2 µL of the

Fe•EDTA solution (1 mM EDTA and 0.5 mM $Fe(NH_4)_2(SO_4)_2.6H_2O$) was added as the catalyst as well as the initiator and the reaction was allowed to proceed for 3-5 min. The reaction was quenched with 10 μL of 100 mM thiourea and concentrated. The DNA was then suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2 min), cooled (0 °C, 5 min), and analyzed by 20% denaturing PAGE.

General Procedure of Hydroxyl Radical Cleavage Reaction for Probing DNA Backbone. The ^{32}P-labeled oligonucleotide (5,000 – 15,000 CPM) was suspended in 8 μL of water and mixed with 10 μL of the 2 × oxidation buffer (20 mM NaCl, 20 mM sodium phosphate, pH 7.2, 2 mM sodium ascorbate, and 0.5 mM hydrogen peroxide). To the reaction 2 μL of the Fe•EDTA solution (1 mM EDTA and 0.5 mM $Fe(NH_4)_2(SO_4)_26H_2O$) was added as the catalyst, as well as the initiator, and the reaction was allowed to proceed for 3-5 min. The reaction was quenched by addition of 10 μL of 100 mM thiourea and subjected precipitated with 3 M NaOAc (4 μL), 1 μg/μL calf thymus DNA (5 μL), and cold ethanol (115 μL). After concentration, the DNA was suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2 min), cooled (0 °C, 5 min), and were analyzed following PAGE (20% denaturing) using a Storm 840 phosphorimager and Imagequant TL software.

General Procedure for Diethylpyrocarbonate (DEPC) Treatment of Oligonucleotides 155 – 157. The ^{32}P-labeled oligonucleotides (5,000 – 15,000 CPM) were suspended in 25.4 μL of 1.18 × PBS buffer (11.8 mM sodium phosphate, pH 7.2,

118 mM NaCl). The oligonucleotides were then incubated with 1.04 M DEPC (final concentration, diluted from neat 6.8 M DEPC) for 10 min at room temperature. The final reaction volume and buffer concentration were 30 μL and 1 × PBS (10 mM sodium phosphate, pH 7.2, 100 mM NaCl), respectively. The reaction tubes were vortexed at an interval of every 2 min during the incubation. Following incubation with DEPC, the samples were precipitated with 5 M NH_4OAc (5.2 μL), 1 μg/μL calf thymus DNA (5 μL), water (10 μL), and cold ethanol (155 μL). After precipitation and concentration, the samples were treated with 1 M piperidine (50 μL) for 20 min at 90 °C, and dried in vacuo. The samples were suspended in water (50 μL) and again dried in vacuo. The DNA was suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2 min), cooled (0 °C, 5 min), and were analyzed following PAGE (20% denaturing) using a Storm 840 phosphorimager and Imagequant TL software.

General Procedure for DEPC Treatment of Oligonucleotides 165 – 167, 135 and 170. The ^{32}P-labeled oligonucleotides (5,000 – 15,000 CPM) were suspended in 25.5 μL of 1.4 × PBS buffer (14 mM sodium phosphate, pH 7.2, 140 mM NaCl). The oligonucleotides were then incubated with 1.75 M DEPC (final concentration, diluted from neat 6.8 M DEPC) for 10 min at room temperature. The final reaction volume and buffer concentration were 35.5 μL and 1 × PBS (10 mM sodium phosphate, pH 7.2, 100 mM NaCl), respectively. The reaction tubes were vortexed at an interval of every 2 min during the incubation. Following incubation with DEPC, the samples were precipitated with 5 M NH_4OAc (5.2 μL), 1 μg/μL calf thymus DNA (5 μL), water (7 μL), and cold ethanol (155 μL). After precipitation and concentration, the samples were treated with 1

M piperidine (50 μL) for 20 min at 90 °C, and dried in vacuo. The samples were suspended in water (50 μL) and again dried in vacuo. The DNA was suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2 min), cooled (0 °C, 5 min), and were analyzed following PAGE (20% denaturing) using a Storm 840 phosphorimager and Imagequant TL software.

General Procedure for $KMnO_4$ Treatment of Oligonucleotides. The ^{32}P-labeled oligonucleotides (5,000 – 15,000 CPM) were suspended in 18 μL of 1.11 × PBS buffer (11.1 mM sodium phosphate, pH 7.2, 111 mM NaCl). For **155 – 158**, the oligonucleotides were then incubated with 0.3 mM $KMnO_4$ (diluted from an aqueous stock solution of 3 mM) for 10 min at room temperature. For **165 – 167**, **135** and **170**, the oligonucleotides were incubated with 0.6 mM $KMnO_4$ (diluted from an aqueous stock solution of 6 mM) for 10 min at room temperature. The final reaction volume and buffer concentration in both instances were 20 μL and 1 × PBS (10 mM sodium phosphate, pH 7.2, 100 mM NaCl), respectively. The reaction tubes were vortexed at an interval of every 2 min during the incubation. Following incubation with $KMnO_4$, the reactions were quenched by addition of neat allyl alcohol (1 μL) and vortexing the reaction tube. The samples were precipitated with 5 M NH_4OAc (5.2 μL), 1 μg/μL calf thymus DNA (5 μL), water (20.5 μL), and cold ethanol (155 μL). After precipitation and concentration, the samples were treated with 1 M piperidine (50 μL) for 20 min at 90 °C, and dried in vacuo. The samples were suspended in water (50 μL) and again dried in vacuo. The oligonucleotides were suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2

min), cooled (0 °C, 5 min), and analyzed following PAGE (20% denaturing) using a Storm 840 phosphorimager and Imagequant TL software.

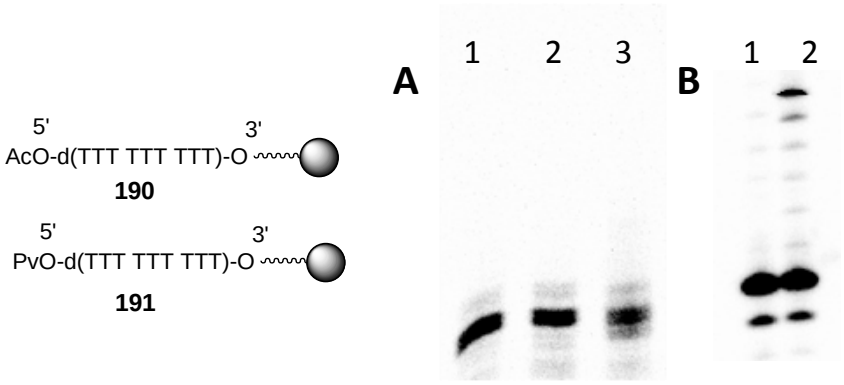
General Procedure for Dimethyl Sulfate (DMS) Treatment of Oligonucleotides. The ^{32}P-labeled oligonucleotides (5,000 – 15,000 CPM) were suspended in 18 μL of 1.11 × PBS buffer (11.1 mM sodium phosphate, pH 7.2, 111 mM NaCl). The oligonucleotides were then incubated with 25 mM DMS (diluted from a 250 nM DMS solution in ethanol) and incubated for 12 min at room temperature. The final reaction volume and buffer concentration were 20 μL and 1 × PBS (10 mM sodium phosphate, pH 7.2, 100 mM NaCl), respectively. The reaction tube was vortexed at an interval of every 2 min during the incubation. Following incubation, the samples were precipitated with 5 M NH_4OAc (5.2 μL), 1 μg/μL calf thymus DNA (5 μL), water (20.5 μL), and cold ethanol (155 μL). After precipitation and concentration, the samples were treated with 1 M piperidine (50 μL) for 20 min at 90 °C, and dried in vacuo. The samples were suspended in water (50 μL) and again dried in vacuo. The oligonucleotides were suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2 min), cooled (0 °C, 5 min), and analyzed following PAGE (20% denaturing) using a Storm 840 phosphorimager and Imagequant TL software.

General Procedure for Preparation of 1.04 M and 2 M Solution of Hydroxylamine (pH 6.0). To make a 10 mL 2 M stock solution of hydroxylamine (pH 6.), 1.34 g hydroxylamine hydrochloride salt was dissolved in 5 mL autoclaved water. Diethylamine (approximately 1.2 mL) was added to adjust the pH to 6.0. Autoclaved

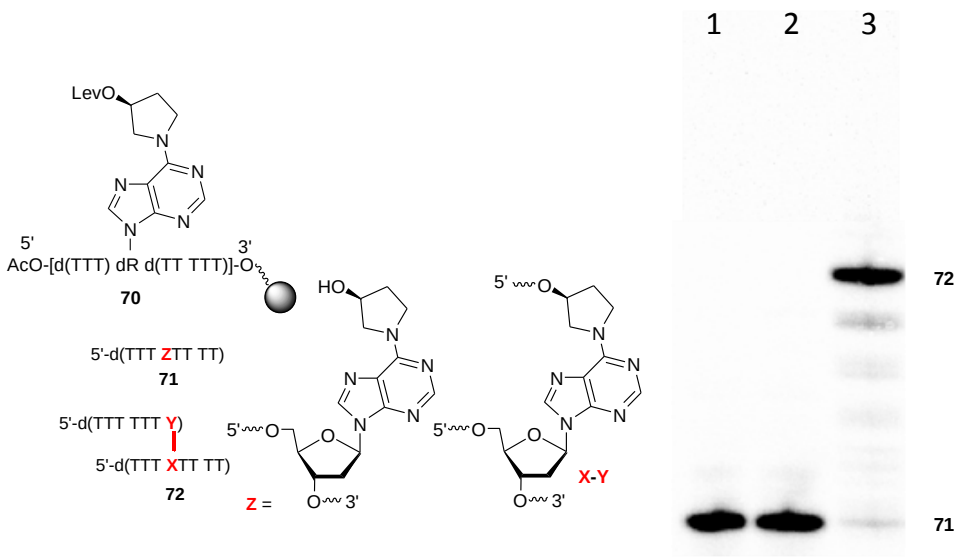
water was added to this solution to achieve a final volume of 10 mL. The stock solution was diluted with autoclaved water to make 1.04 M hydroxylamine solution. The 2 M stock solution can be stored in a -20°C freezer and can be used for one week.

General Procedure for Hydroxylamine Treatment of Oligonucleotides. The ^{32}P-labeled oligonucleotides (5,000 – 15,000 CPM) were suspended in 6 μL of 4.3 × PBS buffer (43.3 mM sodium phosphate, pH 7.2, 433 mM NaCl). The oligonucleotides were then incubated with 0.80 M hydroxylamine, pH 6.0 (diluted from a 1.04 M stock solution) and incubated for 10 min at room temperature. The final reaction volume and buffer concentration were 26 μL and 1 × PBS (10 mM sodium phosphate, pH 7.2, 100 mM NaCl), respectively. The reaction tube was vortexed at an interval of every 2 min during the incubation. Following incubation, the samples were precipitated with 5 M NH_4OAc (5.2 μL), 1 μg/μL calf thymus DNA (5 μL), water (20.5 μL), and cold ethanol (155 μL). After precipitation and concentration, the samples were treated with 1 M piperidine (50 μL) for 20 min at 90 °C, and dried in vacuo. The samples were suspended in water (50 μL) and again dried in vacuo. The oligonucleotides were suspended in formamide loading buffer (10 μL, 50% formamide, 5 mM EDTA, pH 8.0, 0.1% xylene cyanol, and 0.1% bromophenol blue), heated (90 °C, 2 min), cooled (0 °C, 5 min), and analyzed following PAGE (20% denaturing) using a Storm 840 phosphorimager and Imagequant TL software.

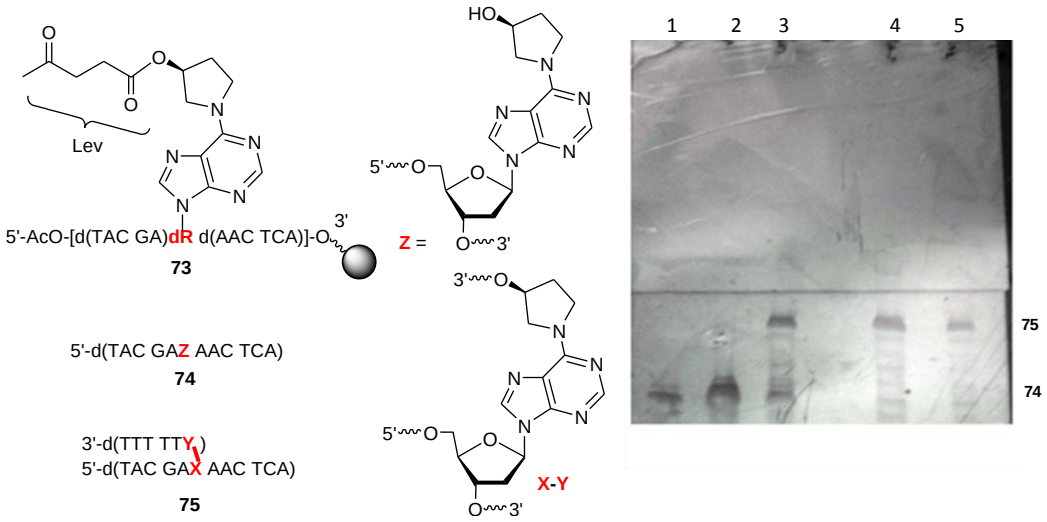
6. Appendix



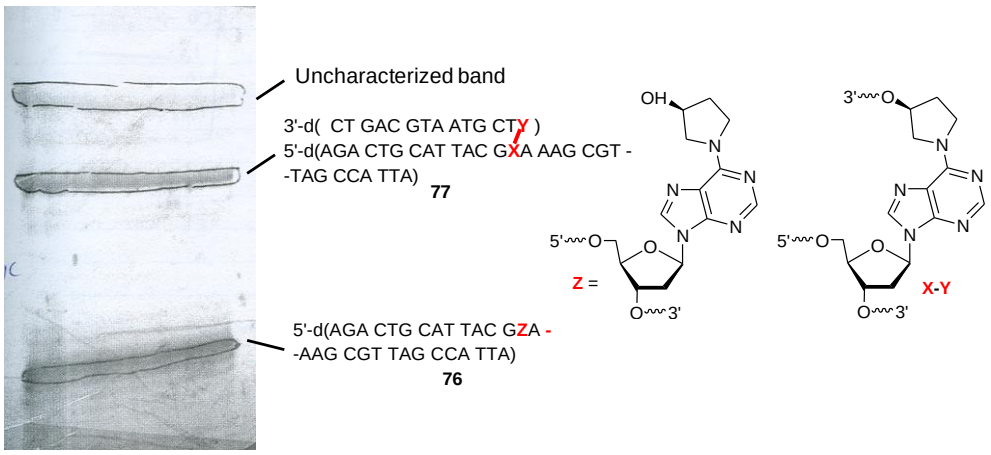
App. Figure 1. Assessment of stability of 5'-O-acetyl and 5'-O-pivaloyl protection on resin bound oligonucleotide under delevulinylation treatment (0.25 M hydrazine in 3:2 pyridine/acetic acid for 10 min, room temperature) after introduction of modified 5'-capping strategy. Delevulinylation and any potential extension of **190** was probed by deprotection, ^{32}P-labeling and analysis of crude radiolabeled materials on 20% denaturing PAGE. (A) Stability of acetyl capping under delevulinylation analyzed on 20% denaturing PAGE Lane 1, **190** deprotected and ^{32}P-labeled. Lane 2, acetyl capped **190** subjected to solid phase extension for 6 dT nucleotides, deprotected and ^{32}P-labeled. Lane 3, acetyl capped **190** subjected to delevulinylation treatment followed by 6 dT extension, deprotected and ^{32}P-labeled. (B) Stability of pivaloyl capping under delevulinylation analyzed on 20% denaturing PAGE. Lane 1, **191** deprotected and ^{32}P-labeled. Lane 2, pivaloyl capped **191** subjected to delevulinylation treatment followed by 6 dT extension, deprotected and ^{32}P-labeled.



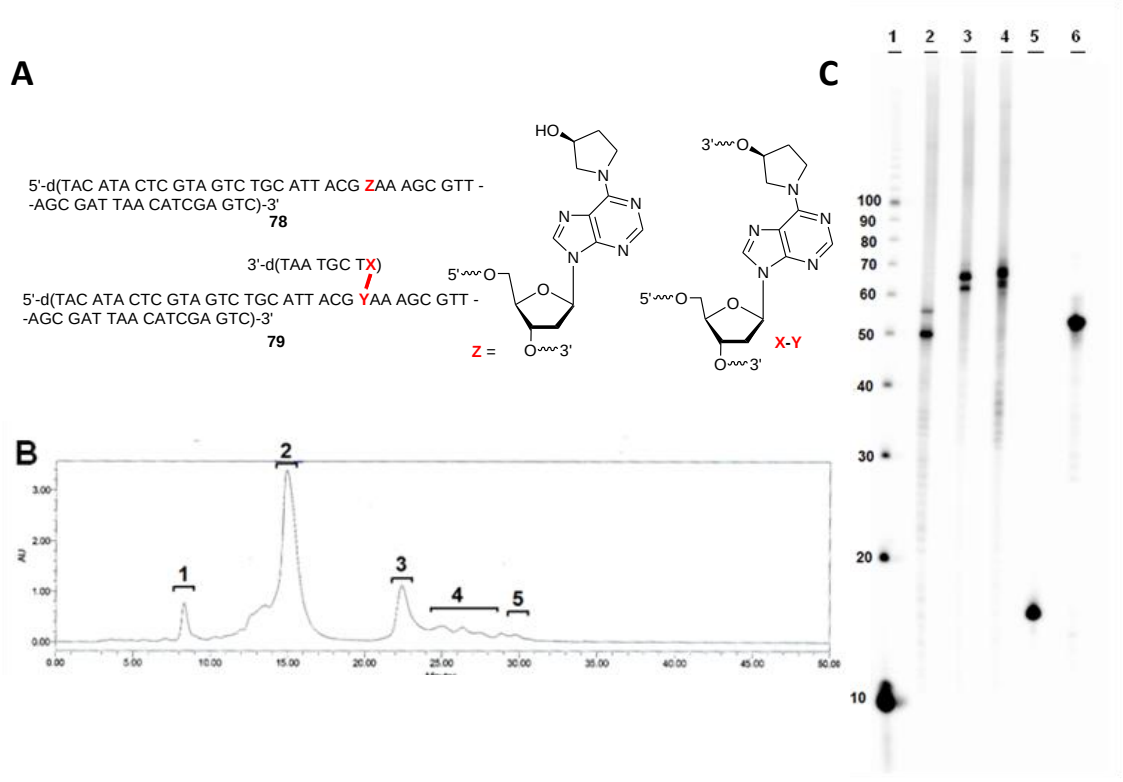
App. Figure 2. Delevulinylation and solid phase extension of **70**. Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Position of **71** (from unmodified **70**) and **72** are annotated. Lane 1, **70** deprotected and radiolabeled. Lane 2, acetyl capped **70** subjected to solid phase extension using 6 dT nucleotides. Lane 3, acetyl capped **70** subjected to delevulinylation treatment followed by 6 dT extension.



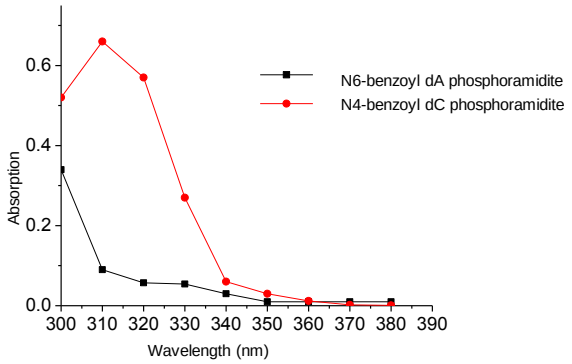
App. Figure 3. Compatibility of commercial nucleobase amine protections under delevulinylation treatment (0.25 M hydrazine in 3:2 pyridine/acetic acid, room temperature) analyzed on a preparative 20% PAGE gel. Benzoyl protection was used for exocyclic amines of A and C while isobutyryl was used for G. The bands are detected under UV lamp with the aid of fluorescent TLC plate. Position of **74** and **75** are annotated. Lane 1, **74** from deprotection of unmodified **73**. Lane 2, **73** after five nucleotide extension using 5'-CE "reverse" T phosphoramidites and ammonolytic deprotection. Lane 3, 4, and 5, **73** subjected to delevulinylation treatment for 1, 5, and 10 min followed by five nucleotide extension using 5'-CE "reverse" T phosphoramidites and ammonolytic deprotection.



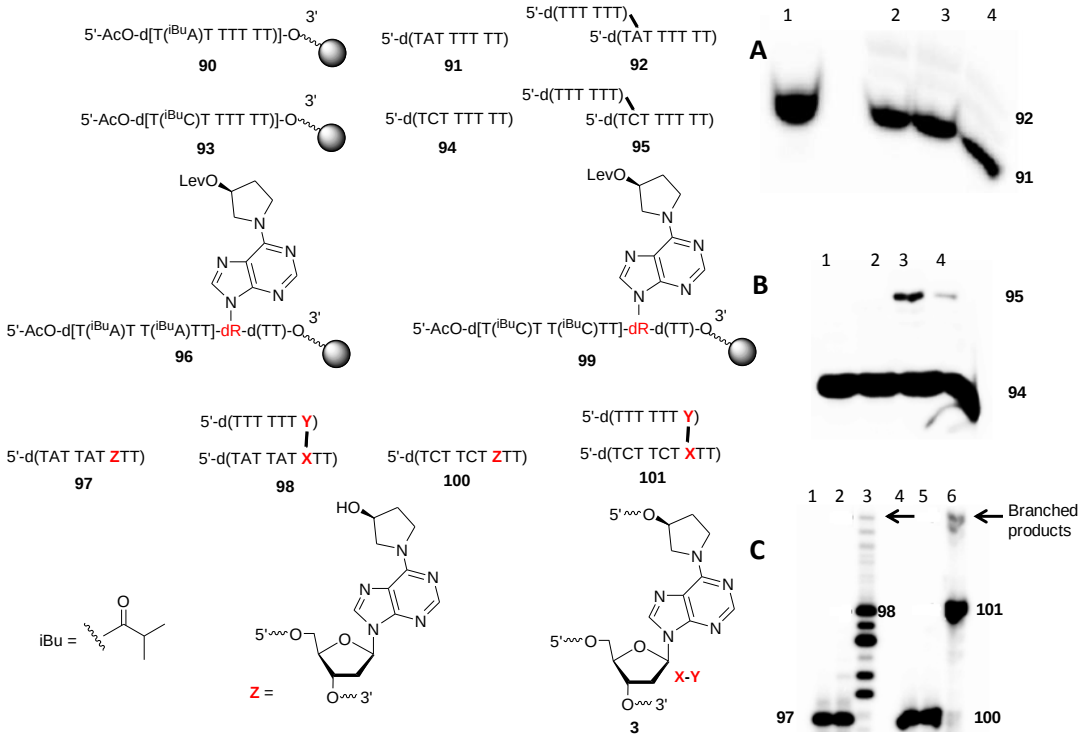
App. Figure 4. Purification of DOB ICL mimic **77**. Following ammonolytic deprotection and cleavage from support, crude oligonucleotide mixture from synthesis of **77** loaded on a preparative 20% denaturing PAGE. The bands are visualized under UV lamp with the aid of a fluorescent TLC plate.



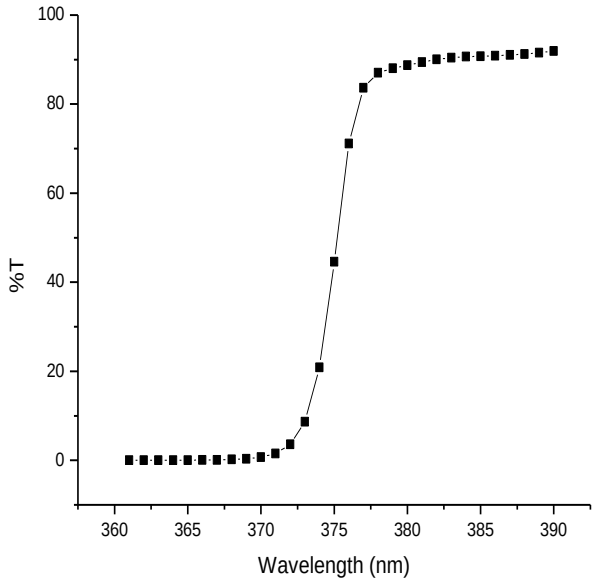
App. Figure 5. HPLC purification of ICL **79**. (A) Template strand **78** and ICL **79**. (B) HPLC chromatogram of purification of **79** in a reverse phase C-18 column. Various isolated fractions (1 – 5) are annotated. (C) Radiolabeling of fractions from the chromatogram of purification of **79** loaded on a 15% denaturing PAGE. Lane 1, 10 bp ladder. Lane 2, fraction 2. Lane 3, fraction 3. Lane 4, fraction 4. Lane 5, ^{32}P-abeled **75**. Lane 6, ^{32}P-labeled **77**.



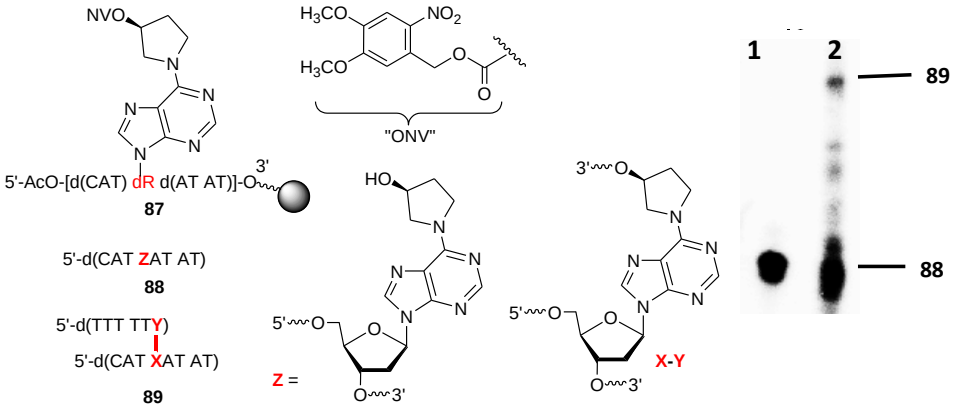
App. Figure 6. Absorbtion of N6-benzoyl-dA and N4-benzoyl dC phosphoramidites (54 μM in acetonitrile).



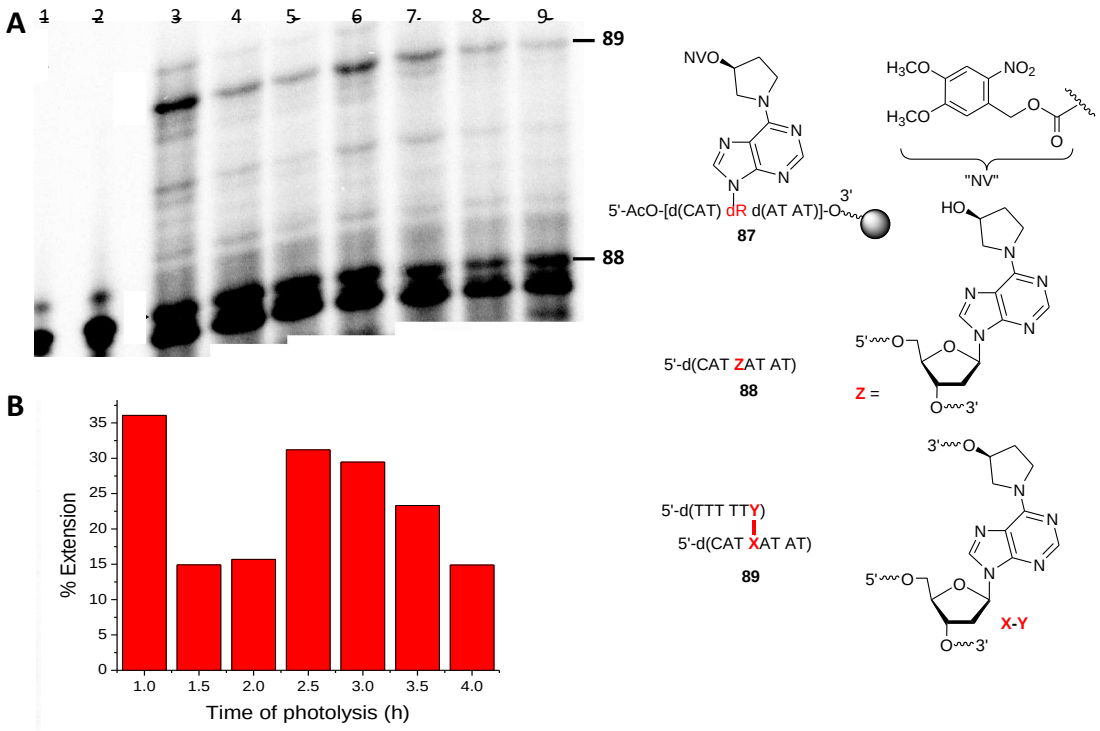
App. Figure 7. Assessment of compatibility of isobutyryl with delevulinylation and subsequent extension for resin bound and 5'-O-acetyl capped oligonucleotide **90**, **93**, **96**, and **99**. Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Positions of template strands and corresponding branching or cross-link product on the gel are annotated. (A) Extension of **90**. Lane 1, **90** deprotected without any modification. Lane 2, **90** subjected to solid phase extension using regular T phosphoramidites. Lane 3 and 4, **90** subjected to delevulinylation and subsequent solid phase extension using regular T phosphoramidites. (B) Extension of **93**. Lane 1, **93** deprotected without any modification. Lane 2, **93** subjected to solid phase extension using regular T phosphoramidites. Lane 3 and 4, **93** subjected to delevulinylation and subsequent solid phase extension using regular T phosphoramidites. (C) Extension of **96** and **99**. Lanes 1 and 4, **96** and **99** deprotected without any modification, respectively. Lanes 2 and 5, **96** and **99** subjected to solid phase extension using regular T phosphoramidites, respectively. Lane 3 and 6, **96** and **99** subjected to delevulinylation followed by solid phase extension using regular T phosphoramidites, respectively.



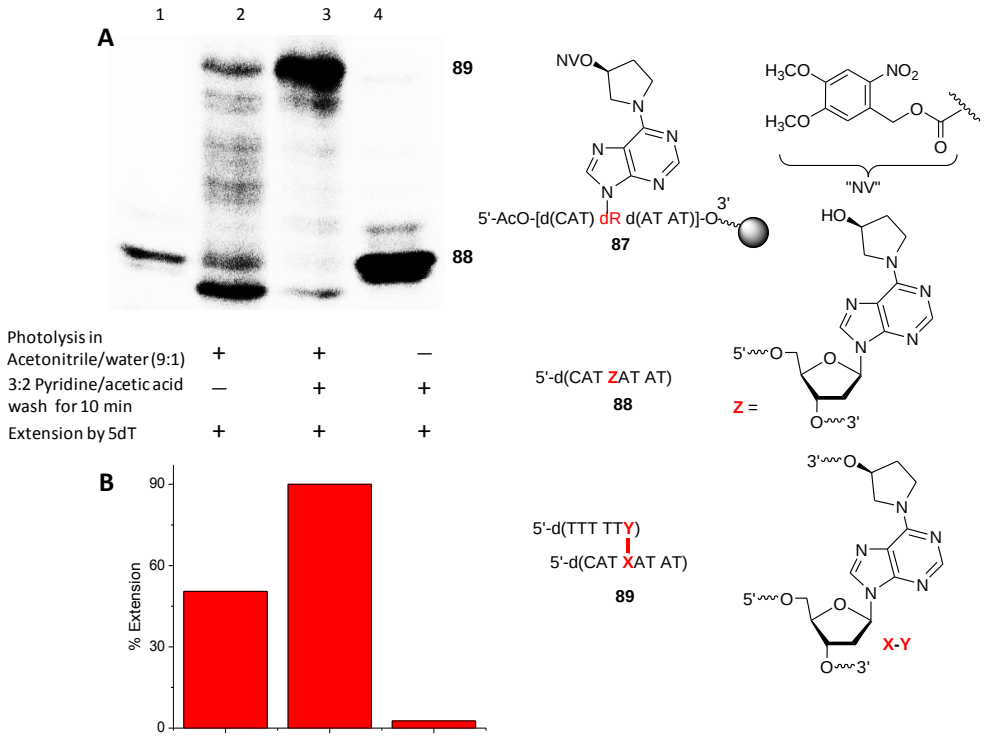
App. Figure 8. Light transmission profile of Newport 375 nm long pass filter (CGA-375).



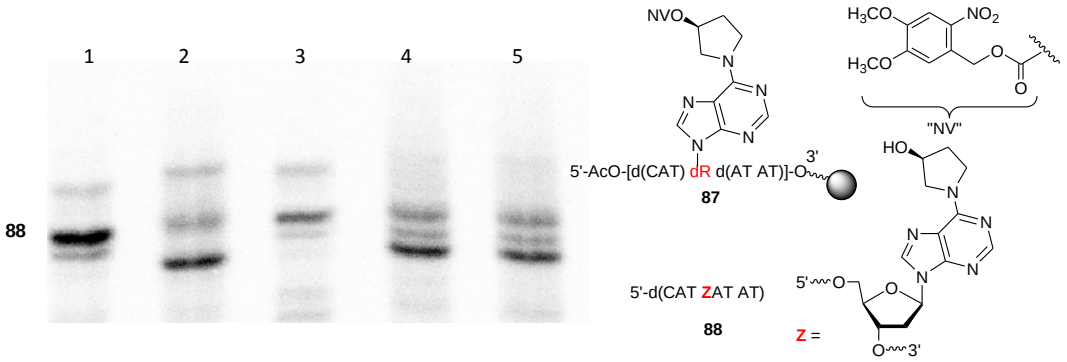
App. Figure 9. Assessment of filter-aided-photolysis (375 nm long pass) and subsequent extension capability for resin bound and 5'-O-acetyl capped oligonucleotide **87**. Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Positions of template strand **88** (from unmodified **87**) and corresponding cross-link product **89** on the gel are annotated. Lane 1, **87** deprotected. Lane 2, **87** photolyzed in toluene for 2 h, extended using regular T phosphoramidites.



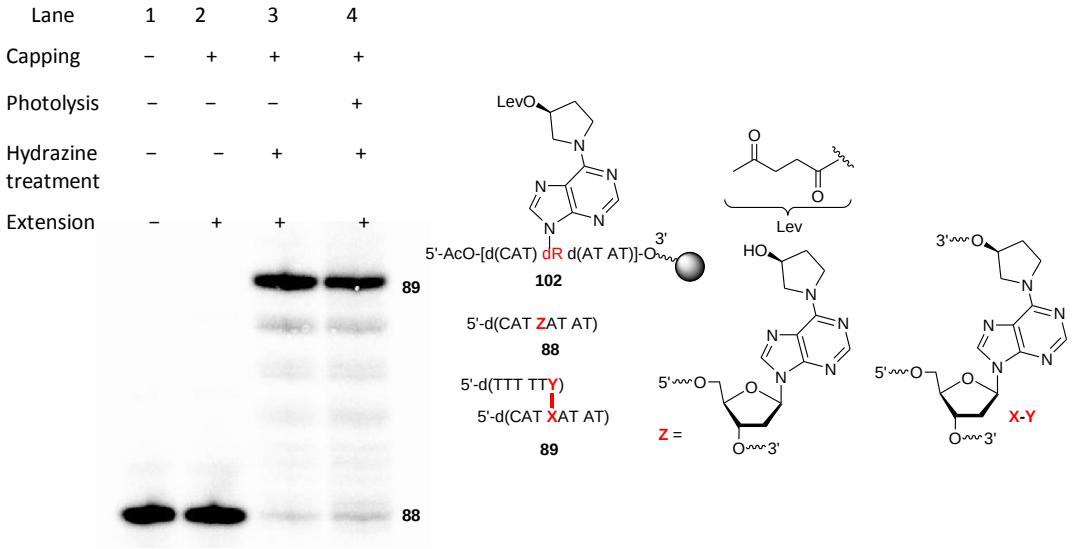
App. Figure 10. Determination of photolysis yield as a function of irradiation time. (A) Photolysis (using 350 nm light source fitted with 375 nm long pass filter) and extension of **87** in toluene. Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Positions of **88** (generated from deprotection of unmodified **87**) and **89** are annotated. Lane 1, unmodified **87**. Lane 2, **87** subjected to extension using regular T phosphoramidites. Lane 3 – 9, **87** subjected to 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0 h of photolysis followed by extension using regular T phosphoramidites. (B) Plot of percentage of extension product obtained with respect to photolysis time. Extension yields were calculated based on the band intensity of template strand and final product.



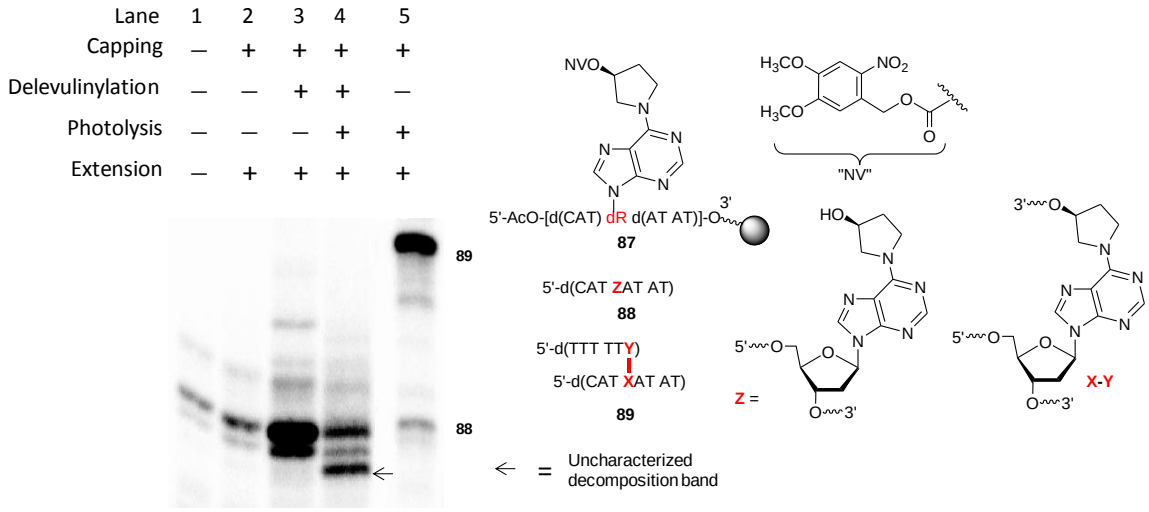
App. Figure 11. Assessment of the effect of 4:1 pyridine/acetic acid washing on photolysis yield. (A) Effect of 4:1 pyridine acetic acid wash on photolysis (using 350 nm light source fitted with 375 nm long pass filter) and subsequent solid phase extension. Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Positions of **88** (generated from deprotection of unmodified **87**) and **89** are annotated. Lane 1, deprotected **87**. Lane 2, **87** subjected to 2.0 h of photolysis in acetonitrile-water (9:1) followed by extension using regular T phosphoramidites. Lane 3, **87** subjected to 2.0 h of photolysis in acetonitrile followed by 3:2 pyridine/acetic acid wash, extension using regular T phosphoramidites. Lane 4, **87** subjected to 2.0 h of photolysis in acetonitrile followed by 3:2 pyridine/acetic acid wash and extension using regular T phosphoramidites. (B) Efficiency of photolysis determined by percentage of **88** converted to **89**. Extension yields were calculated based on the band intensity of template strand and final product.



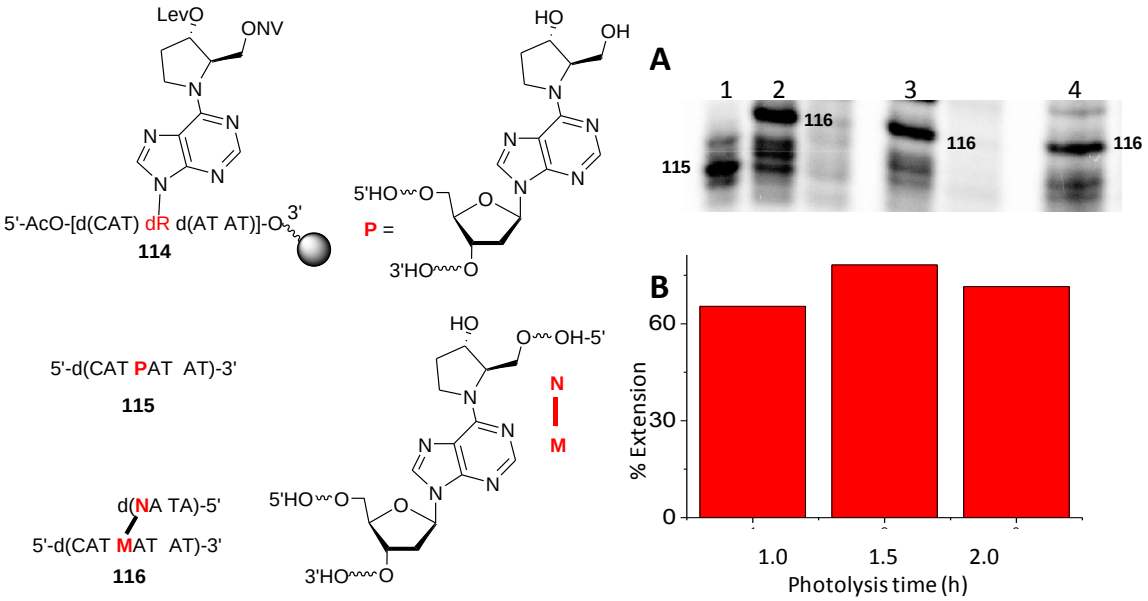
App. Figure 12. Assessment of the effect of 4:1 pyridine/acetic acid wash on photolysis. Photolysis (using 350 nm light source fitted with 375 nm long pass filter) and extension of **87.** Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Position of **88** (generated from unmodified **87**) is annotated. Lane 1, **87** without any modification. Lane 2, photolysis of **87** in 4:1 pyridine/acetic acid for 1 h. Lane 3, resin-bound **87** submerged in 4:1 pyridine/acetic acid for 1 h followed by extension. Lane 4, photolysis of **87** in toluene for 1 h, then washed with 4:1 pyridine/acetic acid. Lane 5, photolysis of **87** in 9:1 acetonitrile/water for 1 h, then washed with 4:1 pyridine/acetic acid.



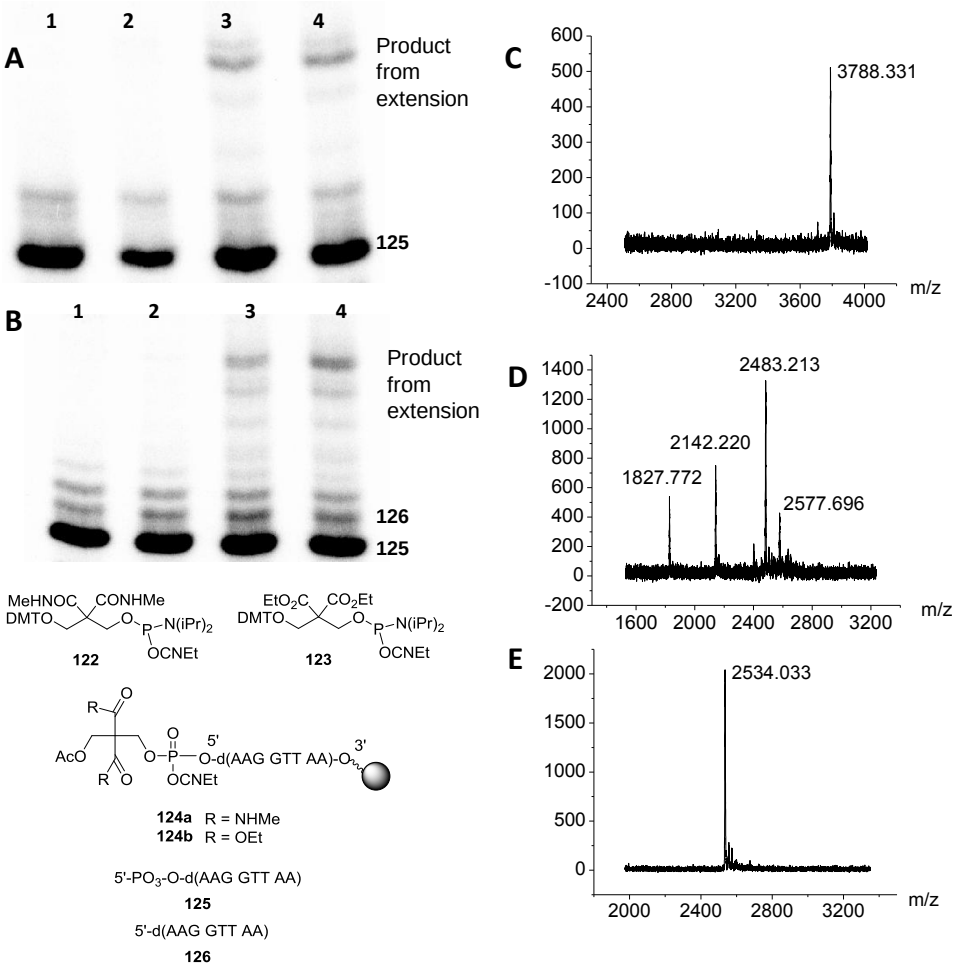
App. Figure 13. Effect of photolysis followed by extension on the solid phase extension of support bound oligonucleotide **102.** Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE. Positions of template strand **88** (from unmodified **87**) and ICL **89** are annotated. Delevulinylation treatment was carried out by treating the resin with 0.25 M hydrazine in 3:2 pyridine/acetic acid for 10 min at room temperature. Photolysis was carried out by irradiating (350 nm light source fitted with 375 nm long pass filter) the resin for 2 h in toluene at room temperature.



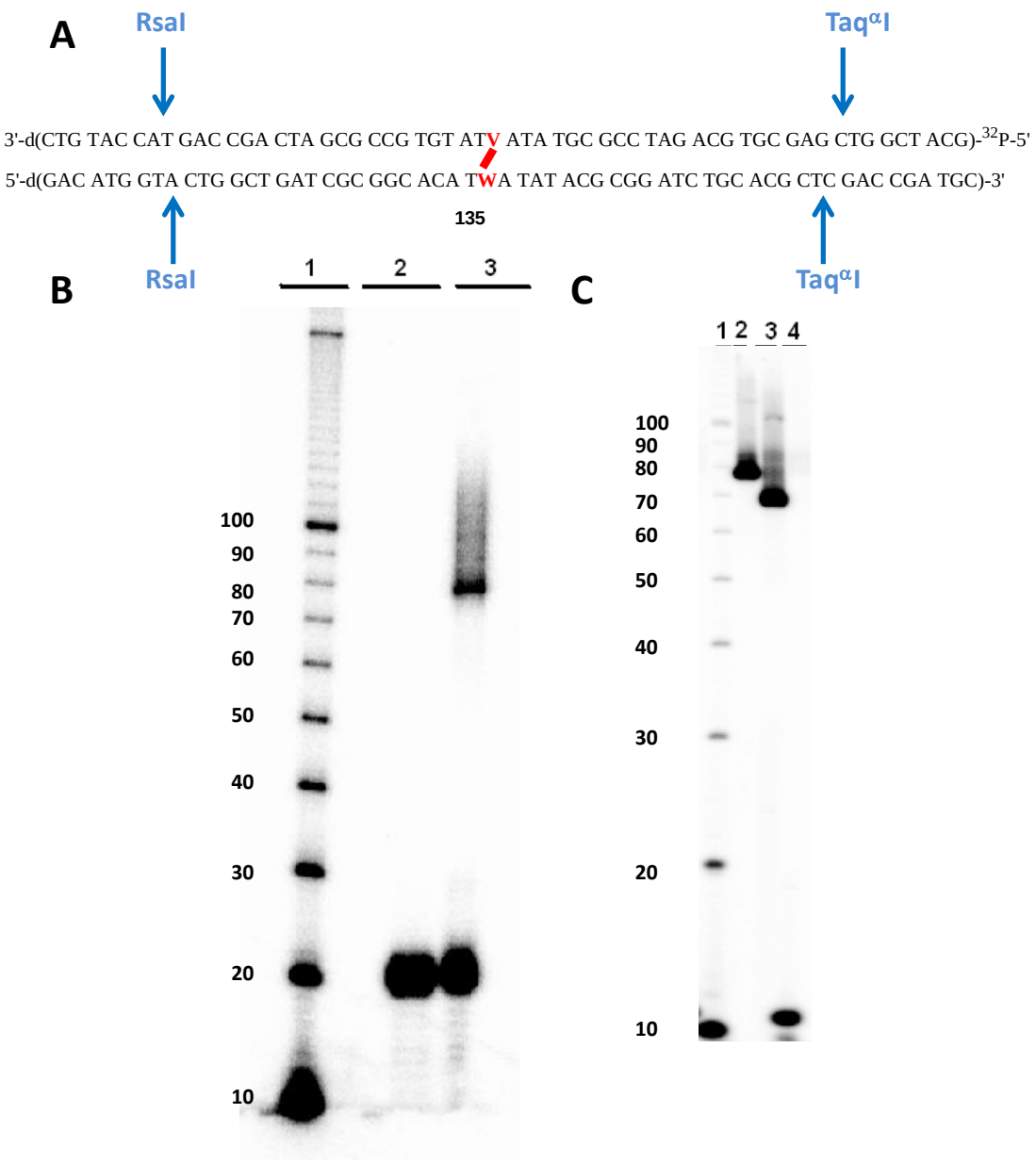
App. Figure 14. Effect of delevulinylation treatment followed by photolysis (using 350 nm light source fitted with 375 nm long pass filter, 2 h, toluene, room temperature) on the solid phase extension of support bound oligonucleotide **87.** Following solid phase synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel. Positions of template strand **88** (from unmodified **87**) and uncharacterized band (presumably from decomposition of **88**) are annotated. Delevulinylation treatment was carried out by treating the resin with 0.25 M hydrazine in 3:2 pyridine/acetic acid for 10 min at room temperature.



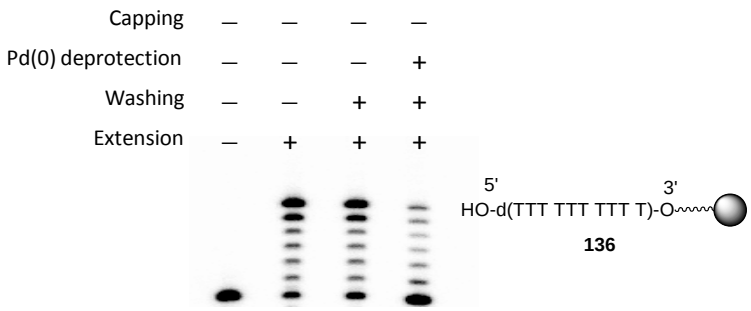
App. Figure 15. Effect of photolysis (using 350 nm light source fitted with 375 nm long pass filter) followed by solid phase extension of support bound oligonucleotide **114.** Following synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), 32P-labeled and analyzed on 20% denaturing PAGE gel. Positions of template strand **115** (from unmodified **114**) and ICL **116** are annotated. (A) Photolysis and 5'-arm extension of **114**. Lane 1, unmodified **114** deprotected. Lane 2, **114** subjected to 1.0 h photolysis in toluene (room temperature) and solid phase extension. Lane 3, **114** subjected to 1.5 h photolysis in toluene (room temperature) and solid phase extension. Lane 4, **114** subjected to 2.0 h photolysis in toluene (room temperature) and solid phase extension. (B) Extension yields plotted against photolysis time. Yields were calculated based on the band intensity of template strand and all extension products.



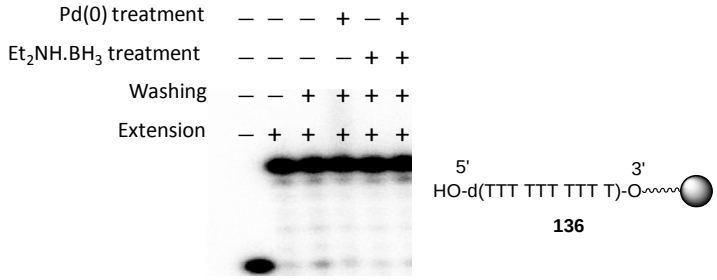
App. Figure 16. Compatibility of phosphorylating agents **122** and **123** with delevulinylation (0.25 M hydrazine in 3:2 pyridine/acetic acid, room temperature for specified time). Following synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), 32P-labeled and analyzed on 20% denaturing PAGE gel. Positions of unphosphorylated template strand **126**, phosphorylated template strand **125** (from unmodified **124a**,**b**) and extension products are annotated. (A) Compatibility of phosphorylating agents **122** with delevulinylation treatment tested using oligonucleotide **124a**. Lane 1, **124a** deprotected. Lane 2, **124a** subjected to solid phase extension using four regular T phosphoramidites. Lane 3, **124a** subjected to delevulinylation treatment for 10 min followed by solid phase extension using regular T phosphoramidites. Lane 4, **124a** subjected to delevulinylation treatment for 25 min followed by solid phase extension using regular T phosphoramidites. (B) Compatibility of phosphorylating agents **123** with delevulinylation treatment tested using oligonucleotide **124b**. Lane 1, **124b** deprotected. Lane 2, **124b** subjected to solid phase extension using four regular T phosphoramidites. Lane 3, **124b** subjected to delevulinylation treatment for 10 min followed by solid phase extension using regular T phosphoramidites. Lane 4, **124b** subjected to delevulinylation treatment for 25 min followed by solid phase extension using regular T phosphoramidites. (C) MALDI-TOF-MS of extension product. (D) MALDI-TOF-MS of **126**. Calc'd mass 2487.689, observed mass 2483.213 $[M + Na]^+$ 2457.689. (E) MALDI-TOF-MS of **125**. Calc'd mass 2436.218, observed mass 2524.022.



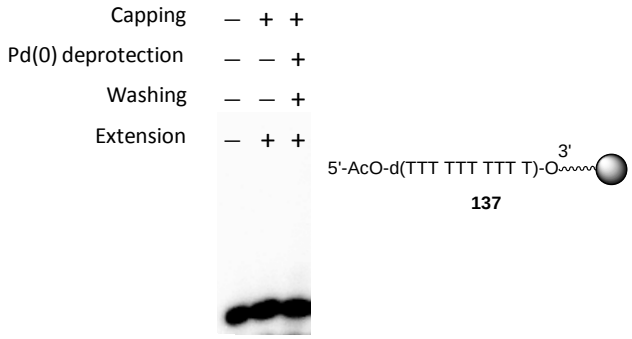
App. Figure 17. Ligation for constructing 5´-^{32}P-**135** and restriction enzyme treatment of purified 5´-^{32}P-**40**. (A) RsaI and Taq$^{\alpha}$I restriction enzyme sites on **135**. (B) Representative 12% denaturing PAGE gel showing the ligation to construct 5´-^{32}P-**135**. Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-**133.** Lane 3, crude reaction to produce 5´-^{32}P-**135.** (C) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of purified 5´-^{32}P-**135**. Lane 1, 10 base pair DNA ladder. Lane 2, 5´-^{32}P-**135** without any treatment. Lane 3, 5´-^{32}P-**135** treated with RsaI. Lane 4, 5´-^{32}P-**135** treated with Taq$^{\alpha}$I.



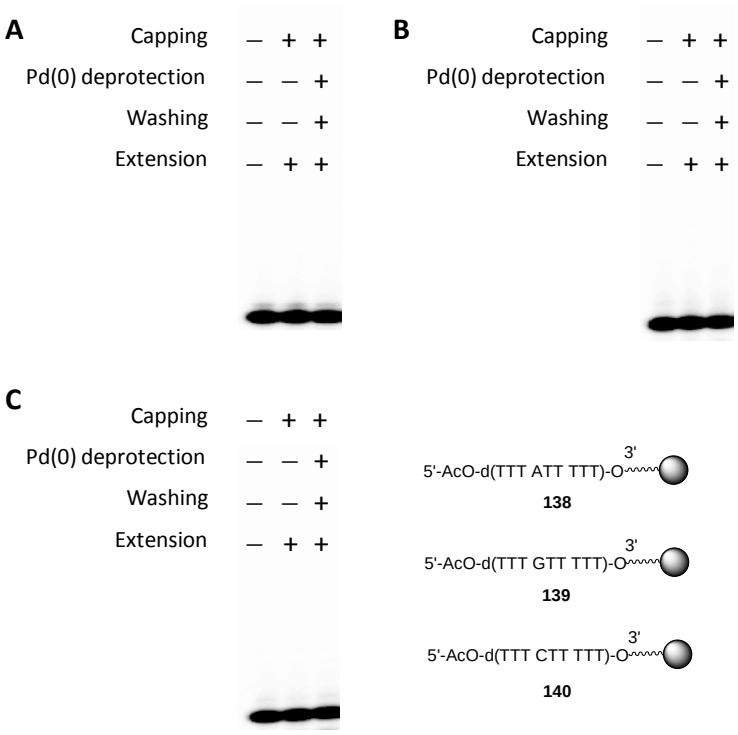
App. Figure 18. Compatibility of solid phase extension with Pd(0) catalyzed deprotection. Support bound **136** was subjected to extension with or without Pd(0) deprotection and washing, followed by six nucleotide extension. Pd(0) deprotection and washing was done as described in "Pd(0) removal of ALLOC: procedure 1"(Experimental Procedure). Following synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel.



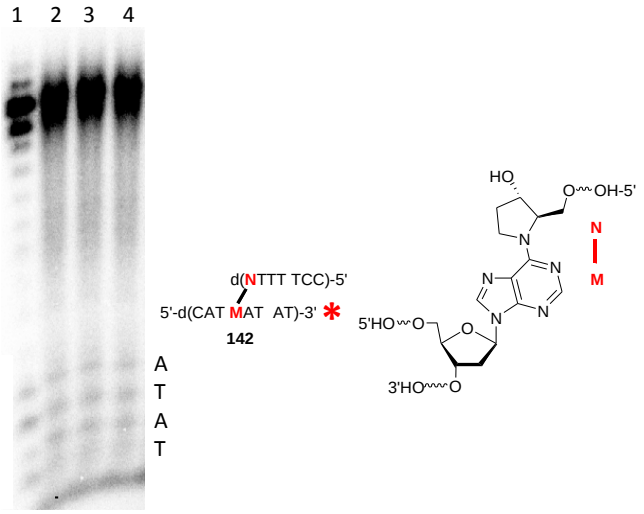
App. Figure 19. Compatibility of solid phase extension with Pd(0) catalyzed deprotection. Support bound **136** was subjected to extension with or without treating the resin with $Pd(PPh_3)4$, $Et_2NH.BH_3$ and washing, followed by six nucleotide extension. Pd(0) deprotection and washing was done as described in "Pd(0) removal of ALLOC: procedure 2"(Experimental Procedure). Following synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel.



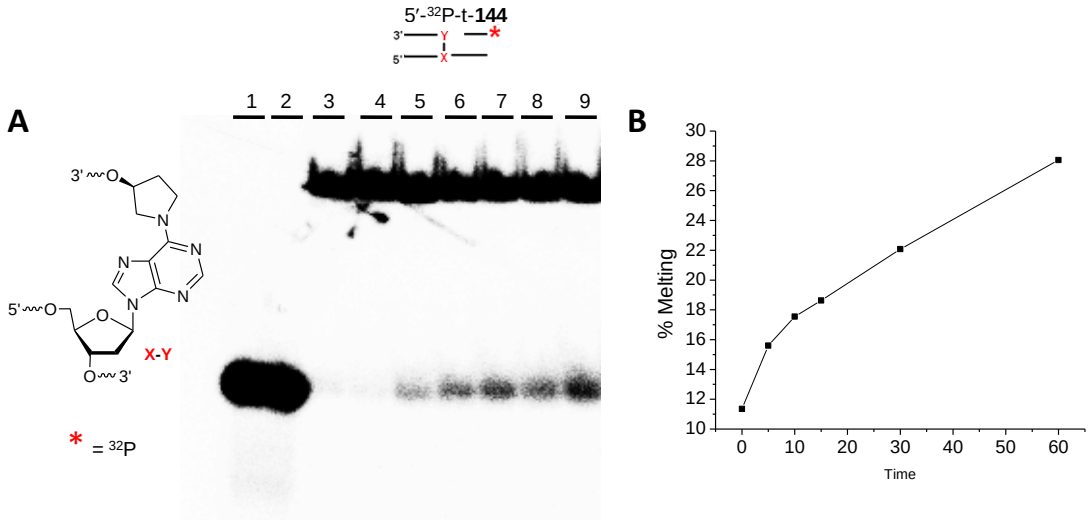
App. Figure 20. Compatibility of 5'-*O*-acyl protection with Pd(0) catalyzed deprotection. Support bound **137** was subjected to extension with or without Pd(0) deprotection and washing, followed by six nucleotide extension. Pd(0) deprotection and washing was done as described in "Pd(0) removal of ALLOC: procedure 2"(Experimental Procedure). Following synthesis, oligonucleotides were deprotected, cleaved from the support (ammonolysis), ^{32}P-labeled and analyzed on 20% denaturing PAGE gel.



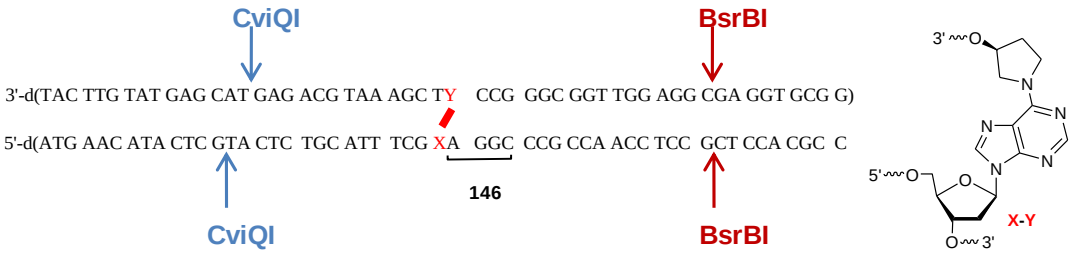
App. Figure 21. Compatibility of exocyclic amine protections on native phosphoramidites [benzoyl-dA (A), isobutyryl-dG (B), and benzoyl-dC (C)] with Pd(0) catalyzed deprotection and washing. Treatment of Support bound and 5'-capped **135** (A), **139** (B), and **140** (C) with or without Pd(0) deprotection and washing, followed by six nucleotide extension. Pd(0) deprotection and washing was done as described in "Pd(0) removal of ALLOC: procedure 2"(Experimental Procedure). Following synthesis, oligonucleotides were deprotected, cleaved from the support by ammonolysis, ^{32}P-labeled and analyzed on 20% denaturing PAGE.



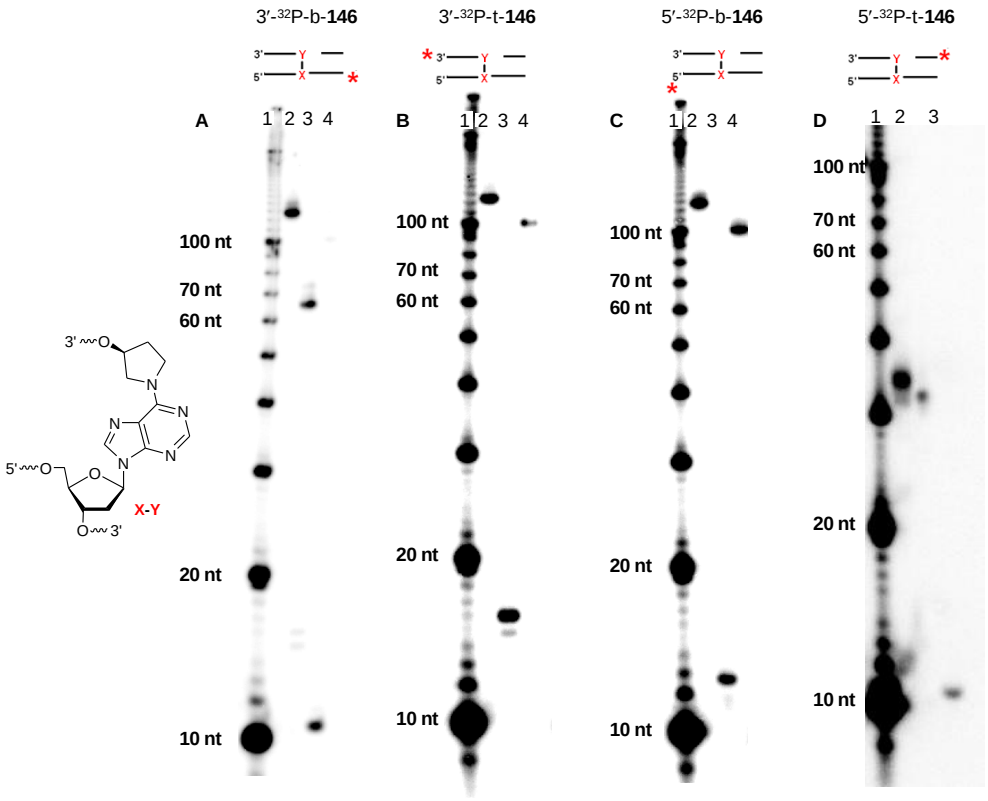
App. Figure 22. Hydroxyl radical cleavage of 3'-^{32}P-**137**. Lane 1, 3'-^{32}P-**137** without any modification. Lane 2, 3 and 4, 3'-^{32}P-**137** subjected to hydroxyl radical cleavage conditions for 3, 4, and 5 min, respectively.



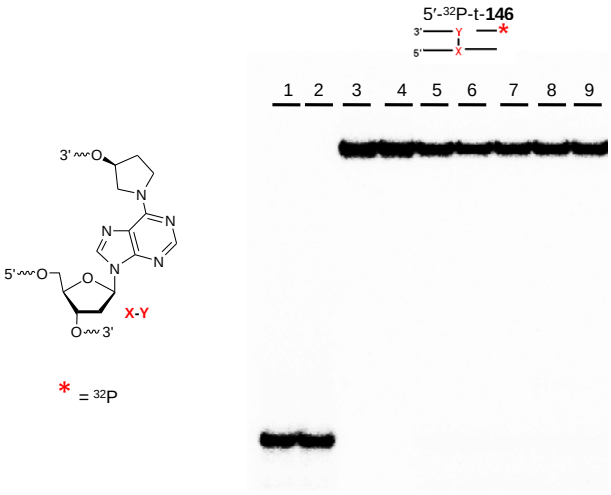
App. Figure 23. Thermal stability of 5'-^{32}P-t-**144** analyzed using non-denaturing (20%) PAGE. (Panel A) Melting of 144 at NER condition [1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM), 55 °C]. Lanes 1 and 2, aliquots at 0 and 60 min from single stranded 5'-^{32}P-**144**, respectively. Lane 3, hybridized 5'-^{32}P-t-**144** without any treatment. Lanes 4 to 9, aliquots taken at 0, 5, 10, 15, 30 and 60 min from 5'-^{32}P-t-**144** (2 nM) incubated with non-radiolabeled **144** (10 nM) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C. (Panel B) Percentage melting plotted against time (min).



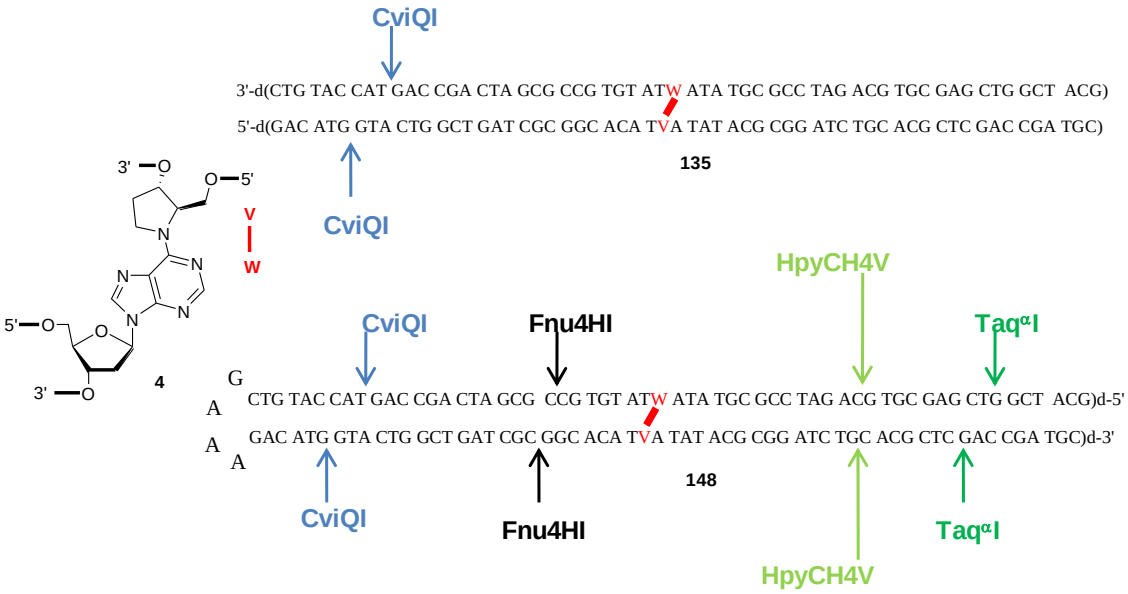
App. Figure 24. Restriction enzyme sites present on **146**.



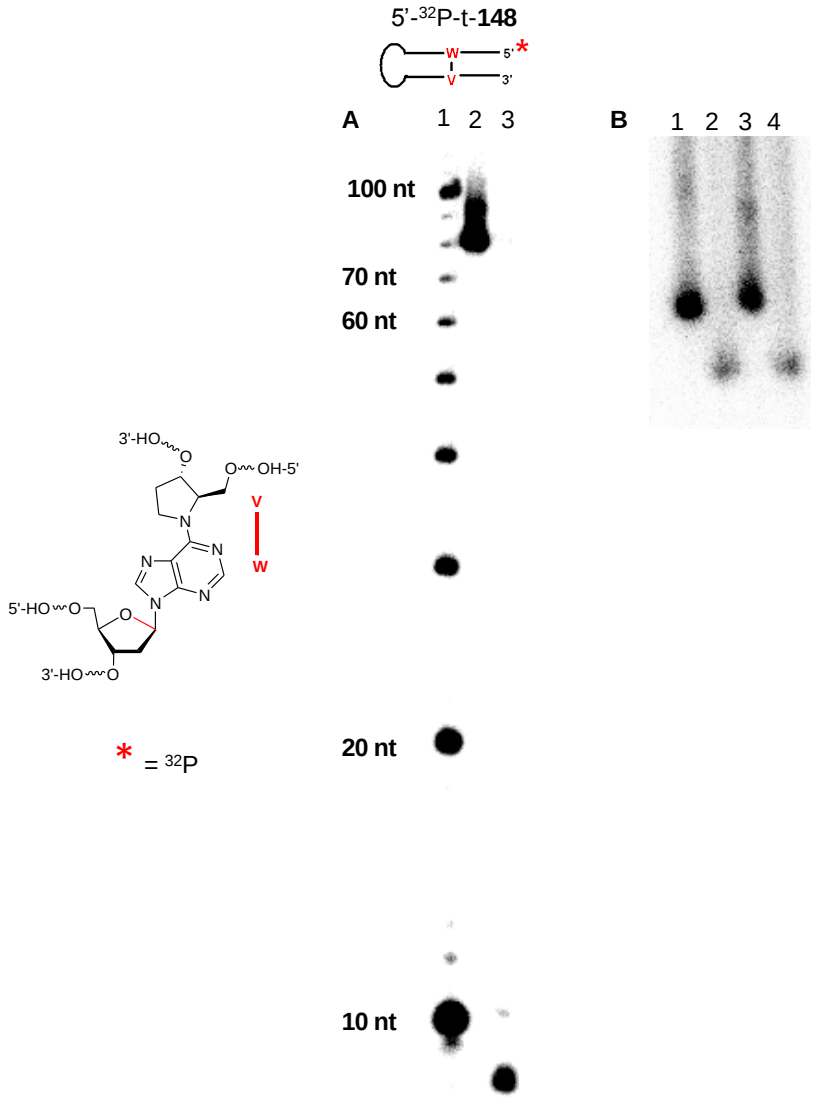
App. Figure 25. Restriction enzyme treatment on 3'-^{32}P-b-**146**, 3'-^{32}P-t-**146**, 5'-^{32}P-b-**146**, and 5'-^{32}P-t-**146**. (A) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of 3'-^{32}P-b-**146**. Lane 1, 10 base pair DNA ladder. Lane 2, 3'-^{32}P-b-**146** without any treatment. Lane 3, 3'-^{32}P-b-**146** treated with CviQI. Lane 4, 3'-^{32}P-b-**146** treated with BsrBI. (B) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of 3'-^{32}P-t-**146**. Lane 1, 10 base pair DNA ladder. Lane 2, 3'-^{32}P-t-**146** without any treatment. Lane 3, 3'-^{32}P-t-**146** treated with CviQI. Lane 4, 3'-^{32}P-t-**146** treated with BsrBI. (C) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of 5'-^{32}P-b-**146**. Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-b-**146** without any treatment. Lane 3, 5'-^{32}P-b-**146** treated with CviQI. Lane 4, 5'-^{32}P-b-**146** treated with BsrBI. (D) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of 5'-^{32}P-t-**146**. Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-t-**146** without any treatment. Lane 3, 5'-^{32}P-t-**146** treated with BsrBI.



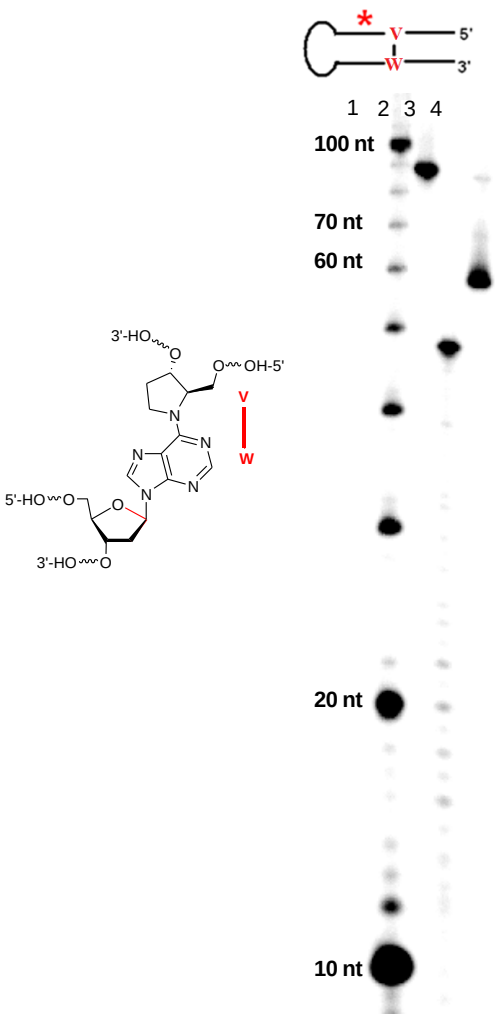
App. Figure 26. Thermal stability of 5'-^{32}P-t-**146** analyzed using non-denaturing (20%) PAGE. Lane 1 and 2, aliquots at 0 and 60 min from single stranded 5'-^{32}P-**146**, respectively. Lane 3, hybridized 5'-^{32}P-t-**146** without any treatment. Lanes 4 to 9, aliquots taken at 0, 5, 10, 15, 30 and 60 min from 5'-^{32}P-t-**146** (2 nM) incubated with non-radiolabeled **146** (10 nM) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C.



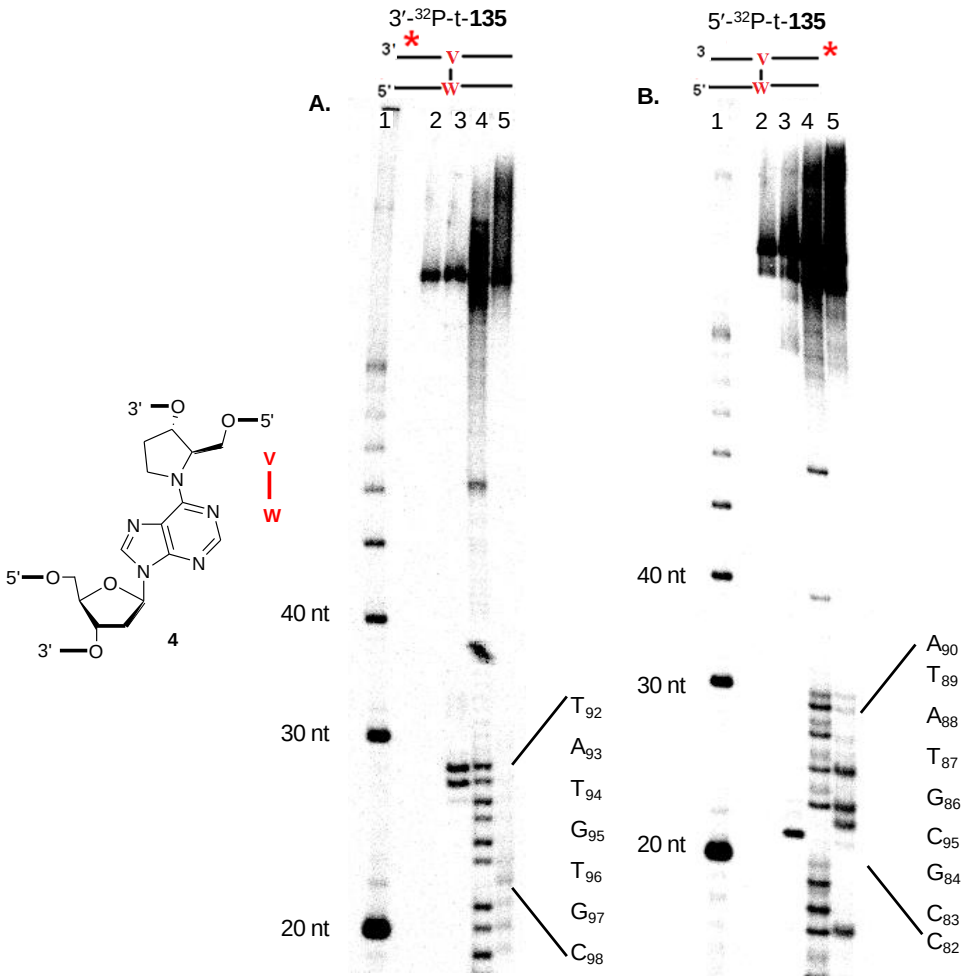
App. Figure 27. Restriction enzyme sites present on **135** and **148.**



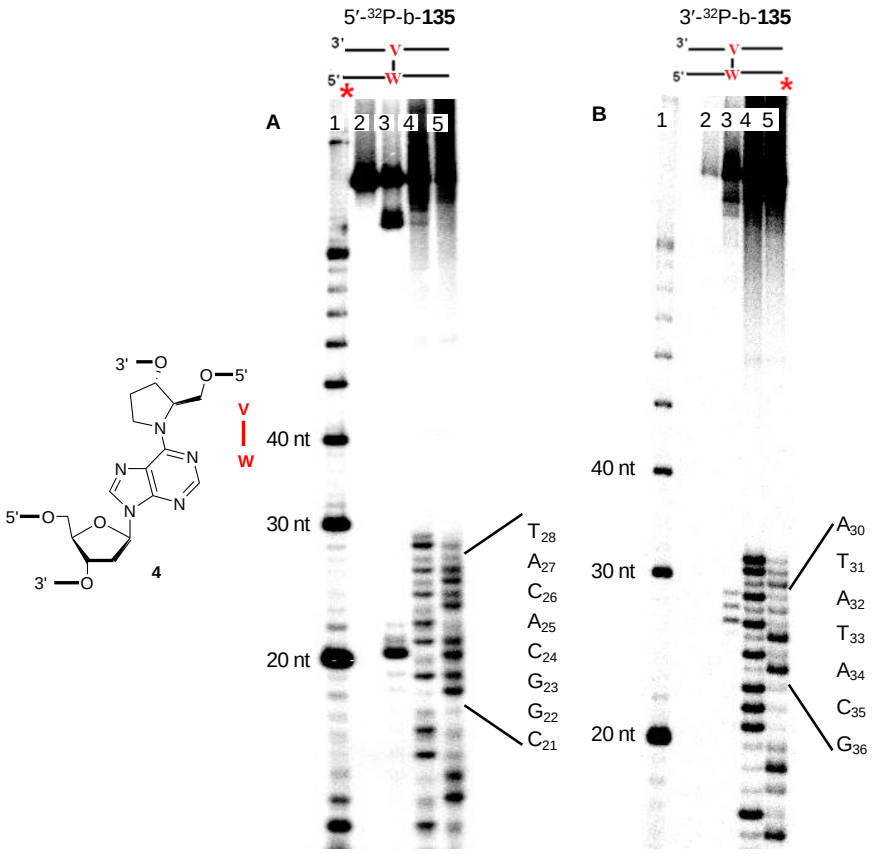
App. Figure 28. Restriction enzyme treatment on 5'-^{32}P-t-**135** and 5'-^{32}P-t-**148.** (A) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of 5'-^{32}P-t-**148**. Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-t-**148** without any treatment. Lane 3, 5'-^{32}P-t-**148** treated with with Taq$^{\alpha}$I. (B) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of 5'-^{32}P-t-**148** and 5'-^{32}P-t-**135**. Lane 1, 5'-^{32}P-t-**135** without any treatment. Lane 2, 5'-^{32}P-t-**135** treated with CviQI. Lane 3, 5'-^{32}P-t-**148** without any treatment. Lane 2, 5'-^{32}P-t-**148** treated with CviQI.



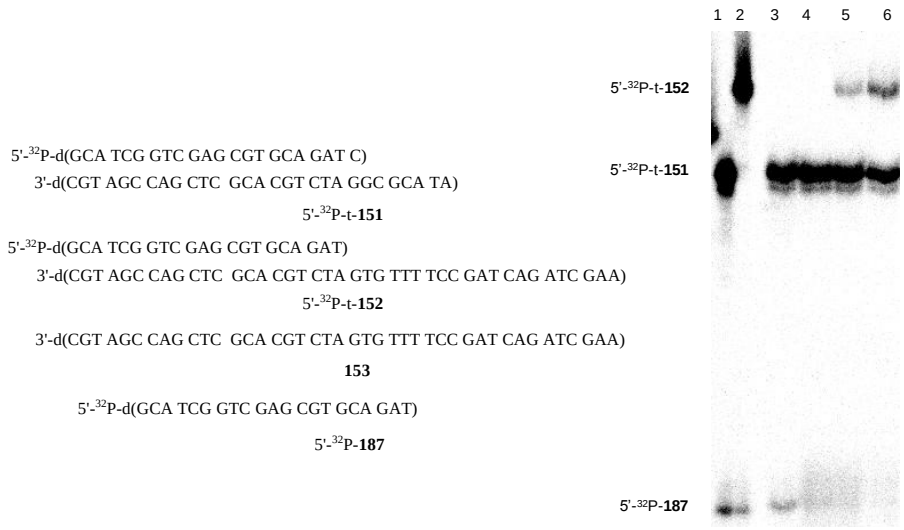
App. Figure 29. Representative 12% denaturing PAGE gel analysis of restriction enzyme treatment on **148** internally labeled at 3'-side of the abasic site analog **(**V**).** Lane 1, 10 base pair DNA ladder. Lane 2, internally labeled **148** without any treatment. Lane 3, internally labeled **148** treated with Fnu4HI. Lane 3, internally labeled 145 treated with HpyCH4V.



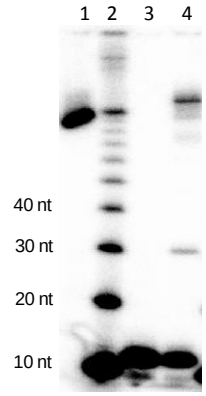
App. Figure 30. Denaturing (20%) PAGE showing UvrABC incision on the top strand (t) of **135**. (Panel A) Strand containing abasic site analog (V) labeled at its 3'-terminus. (Panel B) Strand containing abasic site analog (V) labeled at its 5'-terminus. Lane 1, 10 base pair DNA ladder. Lane 2, 3'-^{32}P-t-**135** (Panel A) or 5′-^{32}P-t-**135** (Panel B). Lane 3, 3'-^{32}P-t-**135** (Panel A) or 5'-^{32}P-t-**135** (Panel B) treated with UvrABC for 8 h. Lane 4, A+G sequencing of 3'-^{32}P-t-**135** (Panel A) or 5'-^{32}P-t-**135** (Panel B). Lane 5, T+C sequencing of 3'-^{32}P-t-**135** (Panel A) or 5'-^{32}P-t-**135** (Panel B).



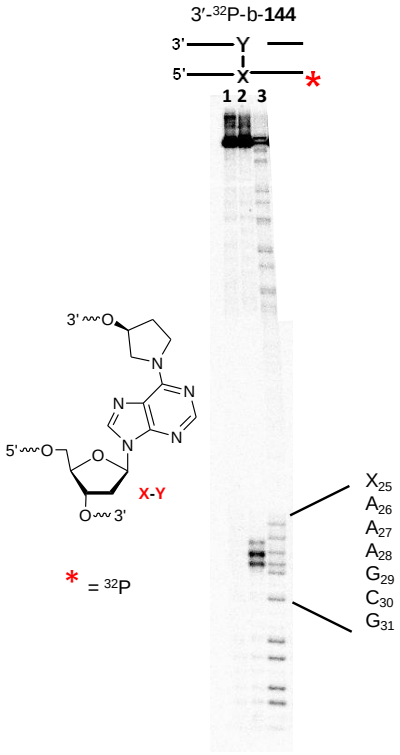
App. Figure 31. Denaturing (20%) PAGE showing UvrABC incision on bottom strand (b) of **135**. (Panel A) Strand containing deoxyadenosine analog (W) labeled at its 5'-terminus. (Panel B) Strand containing deoxyadenosine analog (W) labeled at its 3'-terminus. Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-b-**135** (Panel A) or 3'-^{32}P-b-**135** (Panel B). Lane 3, 5'-^{32}P-b-**135** (Panel A) or 3'-^{32}P-b-**135** (Panel B) treated with UvrABC for 8 h. Lane 4, A+G sequencing of 5'-^{32}P-b-**135** (Panel A) or 3'-^{32}P-b-**135** (Panel B). Lane 5, T+C sequencing of 5'-^{32}P-b-**135** (Panel A) or 3'-^{32}P-b-**135** (Panel B).



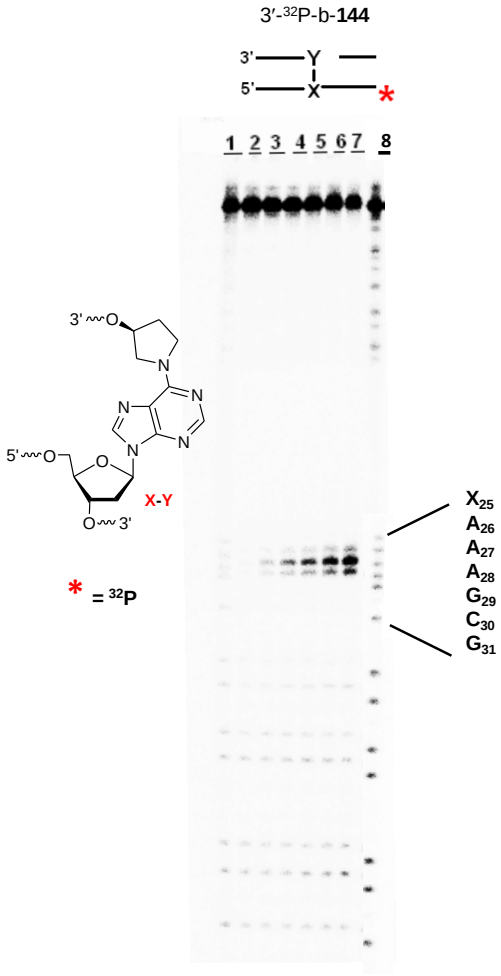
App. Figure 32. Non-denaturing (16%) PAGE analysis of thermal stability of putative UvrABC double strand cleavage product of **135**. Lane 1, 5'-^{32}P-t-**151**. Lane 2, 5'-^{32}P-t-**151**. Lane 3, 5'-^{32}P-t-**151** (0.26 nM) incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at room temperature for 3 h. Lane 4, 5'-^{32}P-t-**151** (0.26 nM) incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 3 h. Lane 5, 5'-^{32}P-t-**151** (0.26 nM) incubated with **153** (0.26 nM) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 3 h. Lane 6, 5'-^{32}P-t-**151** (0.26 nM) incubated with **153** (1.3 nM) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 3 h.



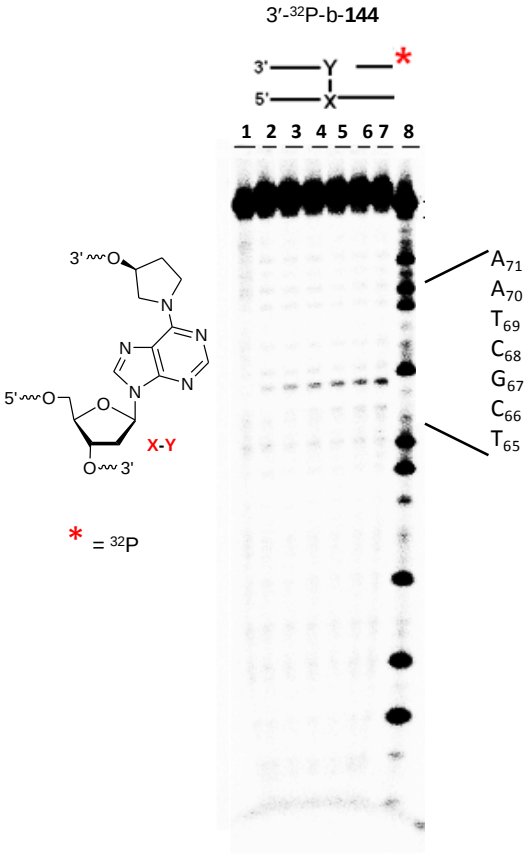
App. Figure 33. Representative 12% gel showing migration of bands from ligation of 3'-^{32}P-b-**189**. Lane 1, 3'-^{32}P-b-**148.** Lane 2, 10 base pair DNA ladder. Lane 3, 3'-^{32}P-**134**. Lane 4, crude reaction to produce 3'-^{32}P-b-**189.**



App. Figure 34. Denaturing PAGE gel (20%) analysis of UvrABC incision on 3'-^{32}P-b-**144**. Lane 1, 3'-^{32}P-b-**144** without any treatment. Lane 2, 3'-^{32}P-b-**144** incubated with UvrABC proteins in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 1 h. Lane 3, A+G sequencing experiment on 3'-^{32}P-b-**144**.

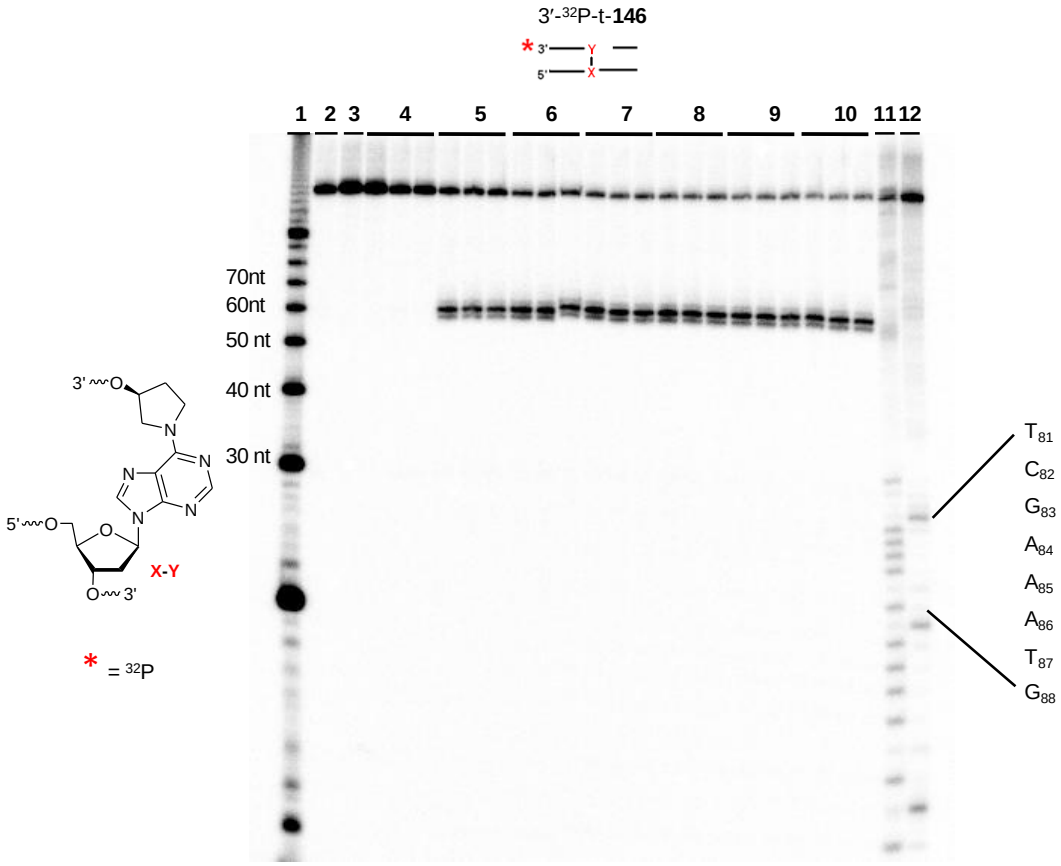


App. Figure 35. Representative denaturing (20%) PAGE showing 1 h time course study of UvrABC incision on 3'-^{32}P-b-**144**. Lane 1, 3'-^{32}P-b-**144** incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) without any treatment at 55 °C for 0 min. Lane 2 to 7, aliquots (in triplicate) taken at 0, 5, 10, 15, 30, and 60 min from incubation of 3'-^{32}P-b-**144** incubated with UvrABC in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C. Lane 8, A+G sequencing reaction on 3'-^{32}P-b-**144**.

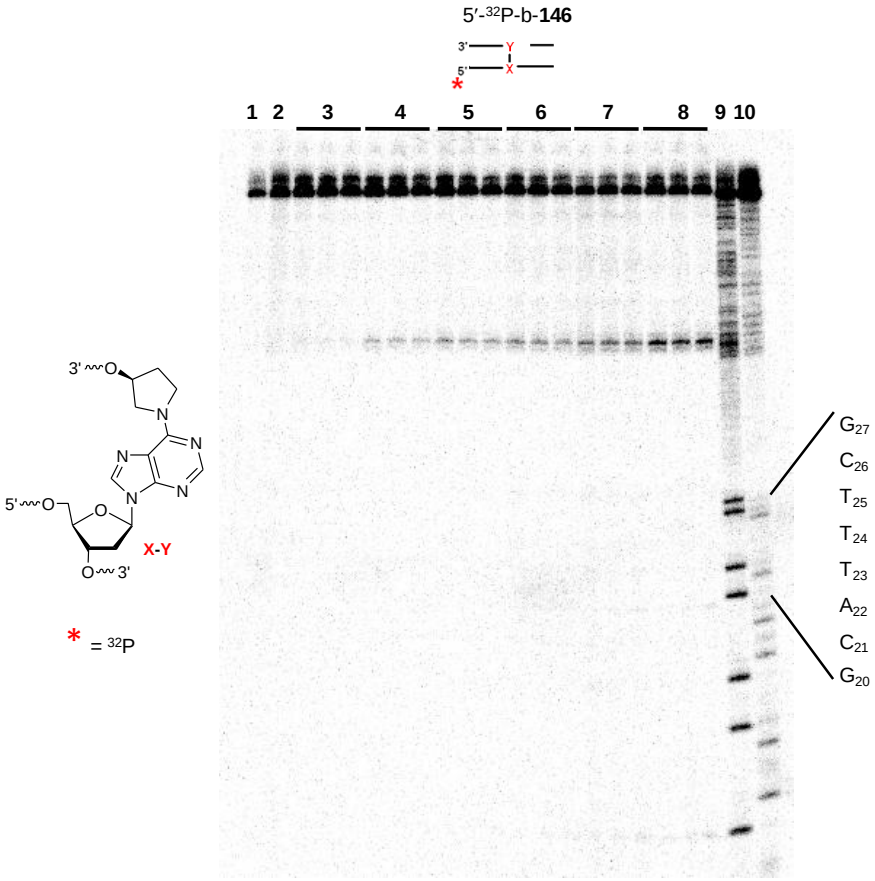


App. Figure 36. Representative denaturing (20%) PAGE showing 1 h time course study of UvrABC incision on 5'-^{32}P-t-**144**. Lane 1, 5'-^{32}P-t-**144** incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) without any treatment at 55 °C for 0 min. Lane 2 to 7, aliquots (in triplicate) taken at 0, 5, 10, 15, 30, and 60 min from incubation of 5'-^{32}P-t-**144** incubated with UvrABC in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C. Lane 8, A+G sequencing reaction on 5'-^{32}P-t-**144**.

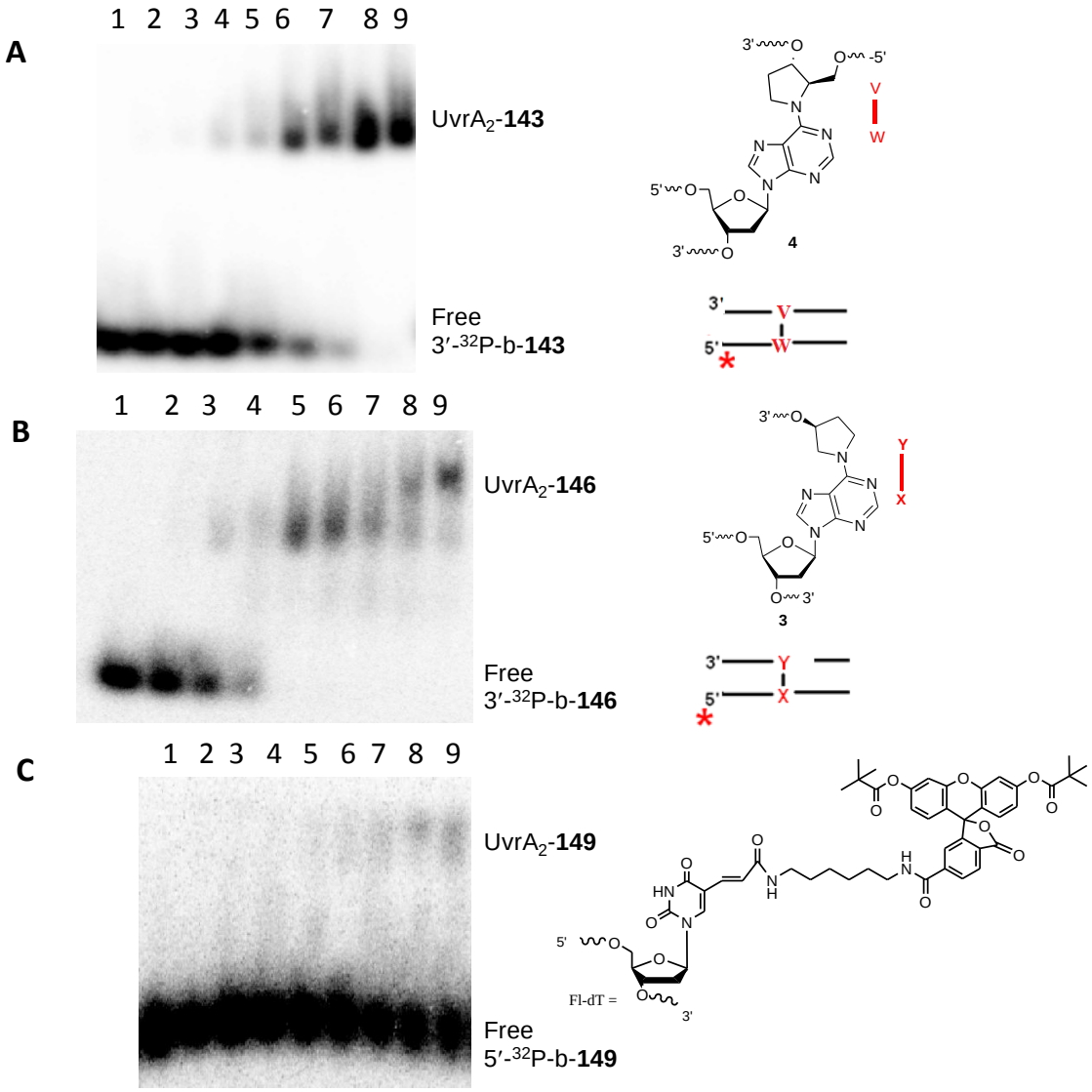
App. Figure 37. Denaturing PAGE gel (20%) analysis of UvrABC incision on 3'-^{32}P-b-**146**. Lane 1, 3'-^{32}P-b-**146** without any treatment. Lane 2, 3'-^{32}P-b-**146** incubated with UvrABC proteins in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 1 h. Lane 3, A+G sequencing experiment on 3'-^{32}P-b-**146**.



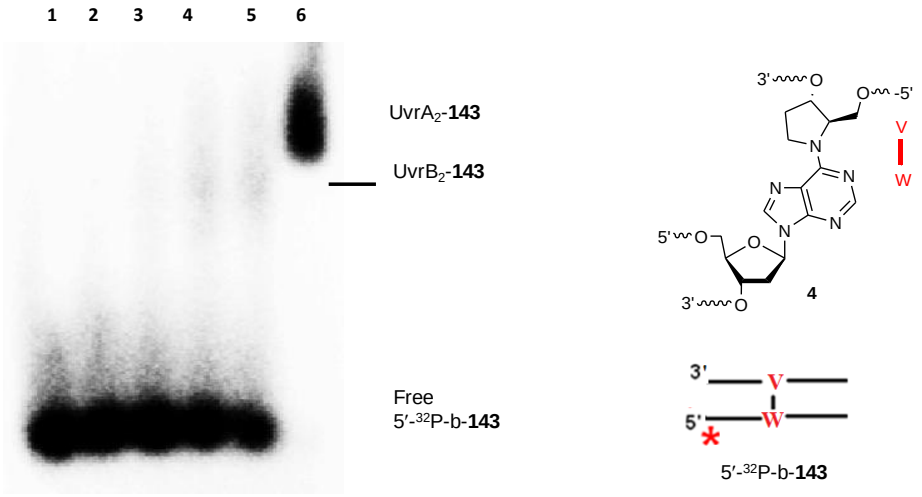
App. Figure 38. Representative denaturing PAGE gel (20%) showing 8 h time course study of UvrABC incision on 3'-^{32}P-t-**146**. Lane 1, 10 base pair DNA ladder. Lane 2, 3'-^{32}P-t-**146** incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) without any treatment at 55 °C for 0 h. Lane 3, 3'-^{32}P-t-**146** incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) without any treatment at 55 °C for 8 h. Lane 4 to 10, aliquots (in triplicate) taken at 0, 1, 2, 3, 4, 6, and 8 h from incubation of 3'-^{32}P-t-**146** incubated with UvrABC in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C. Lane 11, A+G sequencing reaction on 3'-^{32}P-t-**146**. Lane 12, T+C sequencing reaction on 3'-^{32}P-t-**146**.



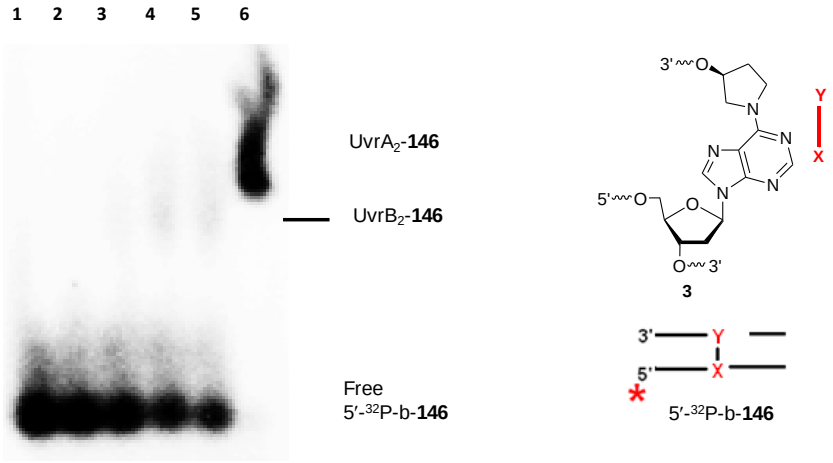
App. Figure 39. Representative denaturing (20%) PAGE showing 1 h time course study of UvrABC incision on 5'-^{32}P-b-**146**. Lane 1, 5'-^{32}P-b-**146** incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) without any treatment at 55 °C for 0 min. Lane 2, 5'-^{32}P-b-**146** incubated in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) without any treatment at 55 °C for 60 min. Lane 3 to 9, aliquots (in triplicate) taken at 0, 5, 10, 15, 30, and 60 min from incubation of 5'-^{32}P-b-**146** incubated with UvrABC in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C. Lane 11, A+G sequencing reaction on 5'-^{32}P-b-**146**. Lane 12, T+C sequencing reaction on 5'-^{32}P-b-**146**.



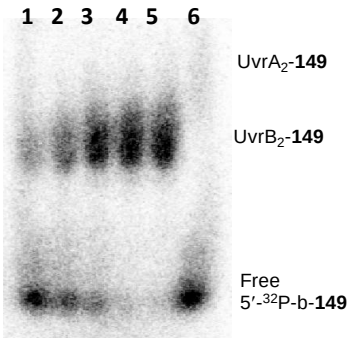
App. Figure 40. UvrA binding of 3'-^{32}P-t-**143**, **3**'-^{32}P-t-**146**, and 5'-^{32}P-t-**149** (containing fluoresceinylated dT or Fl-dT. Aliquots from DNA-protein incubations are analyzed by gel shift assay (4% nondenaturing PAGE). Positions of $UvrA_2$-DNA complex and free DNA are annotated. Lanes 1 – 9, incubation of UvrA (0.5, 5, 10, 20, 40, 60, 80, 120, and 200 nM, respectively) with 0.3 nM 3'-^{32}P-t-**143** (A) or 3'-^{32}P-t-**146** (B) or 5'-^{32}P-t-**149** (C) in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min.



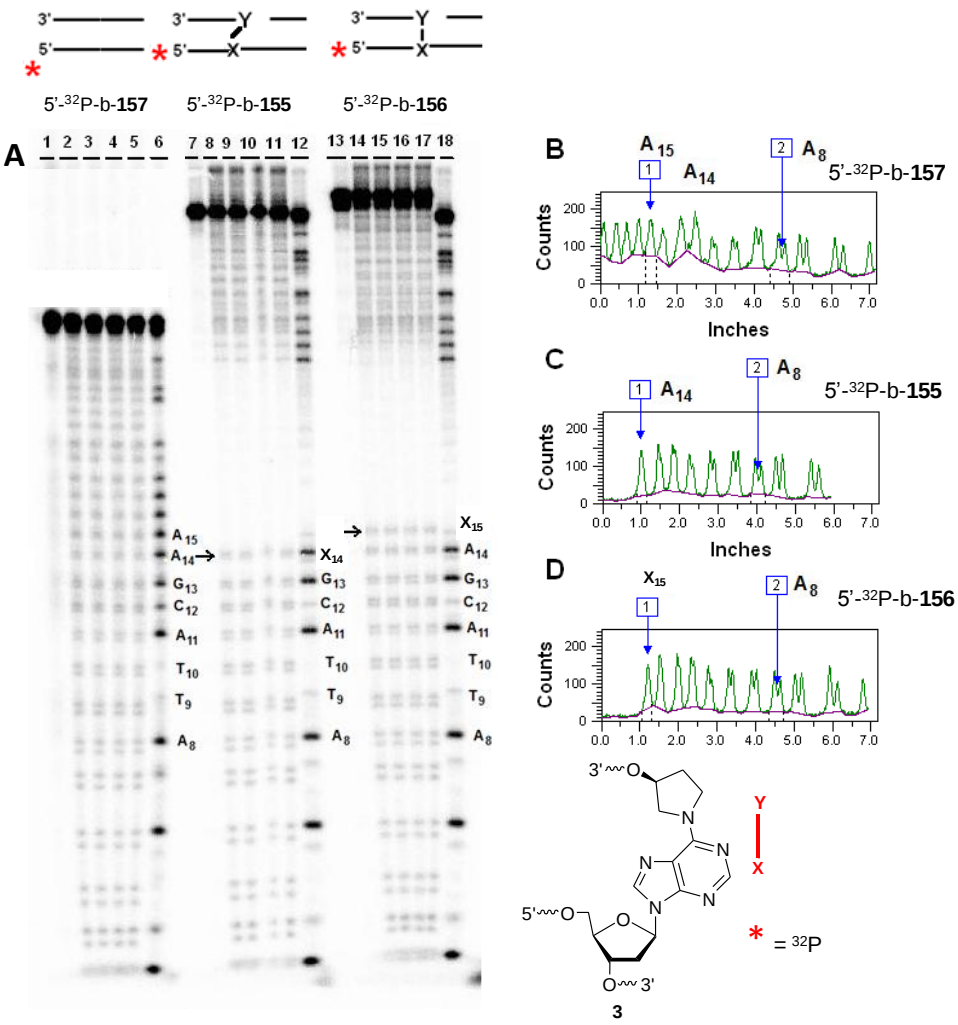
App. Figure 41. UvrA binding and subsequent transfer of 5'-^{32}P-t-**143** to UvrB. Aliquots from the reaction are analyzed by gel shift assay (4% nondenaturing PAGE). Positions of free DNA, $UvrA_2$-DNA, $UvrB_2$-DNA complexes are annotated. Lanes 1 – 5, incubation of UvrA (1, 2, 4, 8 and 10 nM, respectively) and UvrB (500 nM) with 0.3 nM 3'-^{32}P-t-**143** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min. Lane 6, incubation of 200 nM UvrA with 0.3 nM 3'-^{32}P-t-**143** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min.



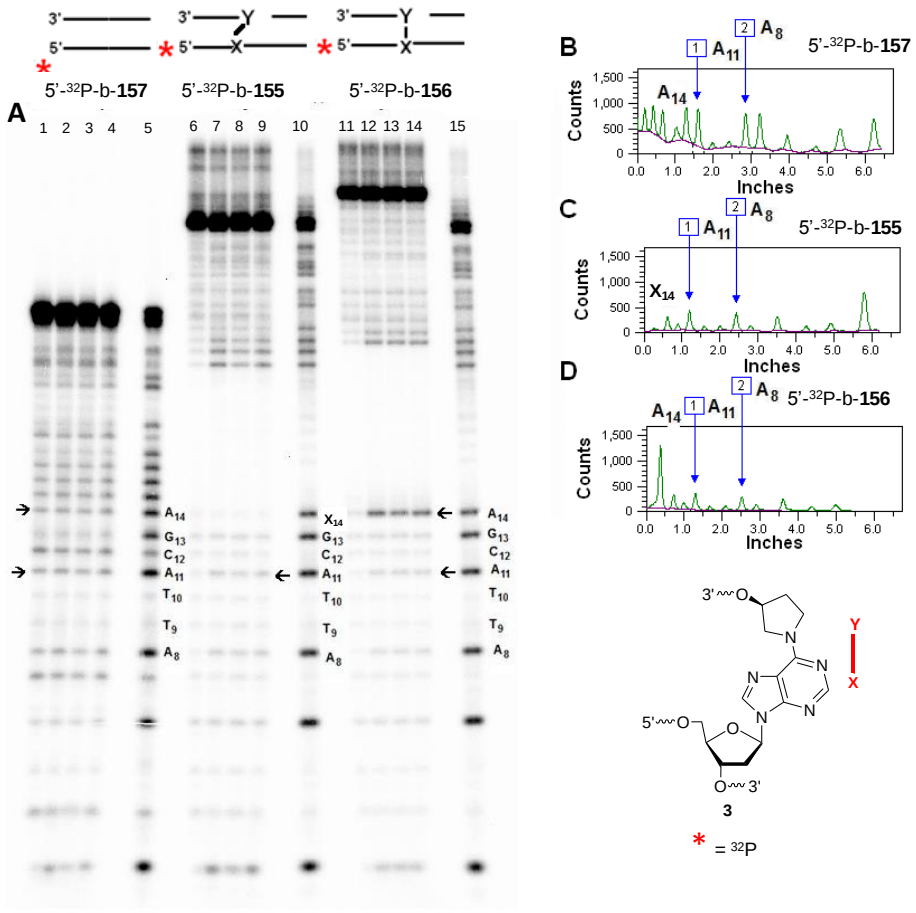
App. Figure 42. UvrA binding and subsequent transfer of 5'-^{32}P-t-**146** to UvrB. Aliquots from the reaction are analyzed by gel shift assay (4% nondenaturing PAGE). Positions of free DNA, $UvrA_2$-DNA, $UvrB_2$-DNA complexes are annotated. Lanes 1 – 5, incubation of UvrA (1, 2, 4, 8 and 10 nM, respectively) and UvrB (500 nM) with 0.3 nM5'-^{32}P-t-**146** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min. Lane 6, incubation of 200 nM UvrA with 0.3 nM 5'-^{32}P-t-**146** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min.



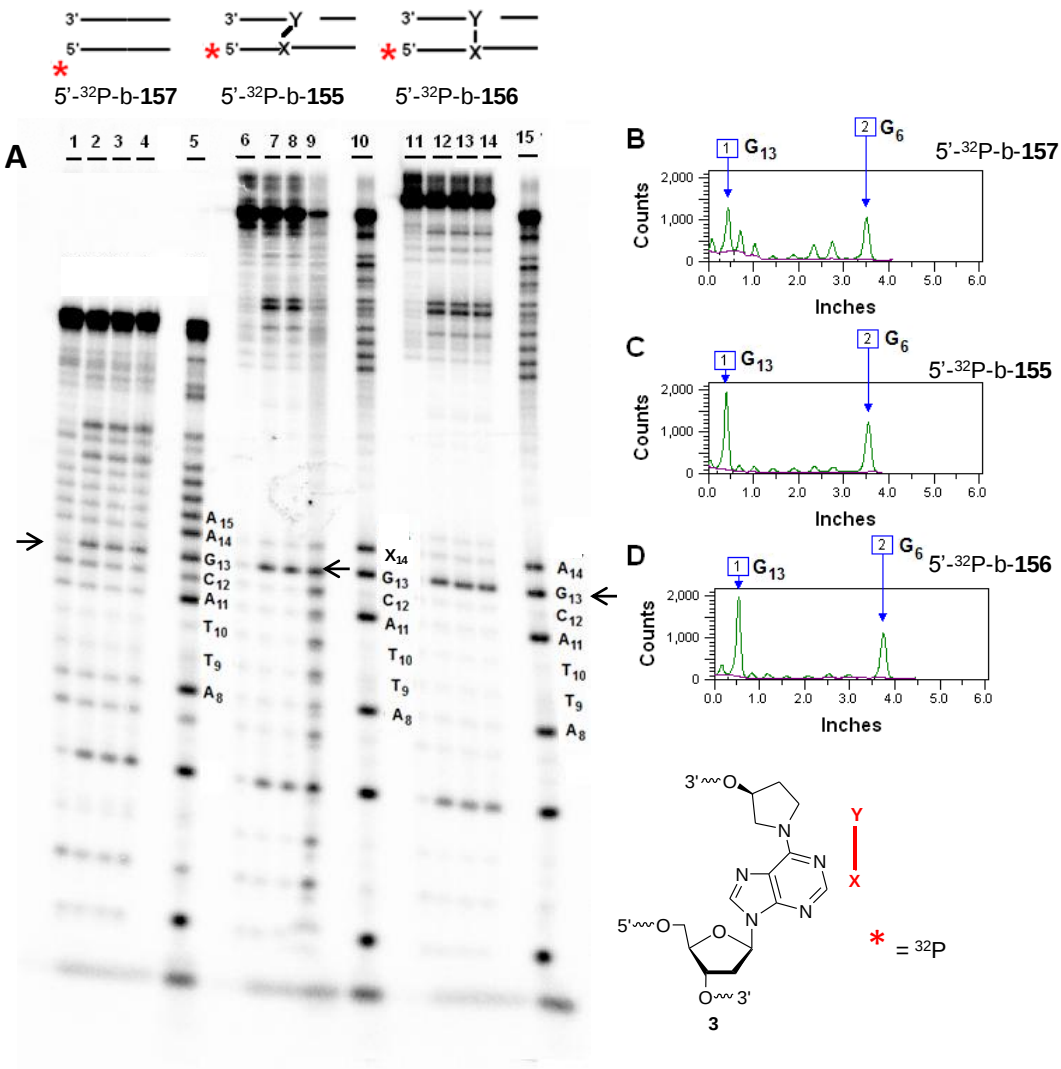
App. Figure 43. UvrA binding and subsequent transfer of 5'-^{32}P-t-**149** to UvrB. Aliquots from the reaction are analyzed by gel shift assay (4% nondenaturing PAGE). Positions of free DNA, $UvrA_2$-DNA, $UvrB_2$-DNA complexes are annotated. Lanes 1 – 5, incubation of UvrA (1, 2, 4, 8 and 10 nM, respectively) and UvrB (500 nM) with 0.3 nM 5'-^{32}P-t-**149** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min. Lane 6, incubation of 200 nM UvrA with 0.3 nM 5'-^{32}P-t-**149** in 1 × NER buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 50 mM KCl, 5 mM dTT, and ATP 1 mM) at 55 °C for 20 min.



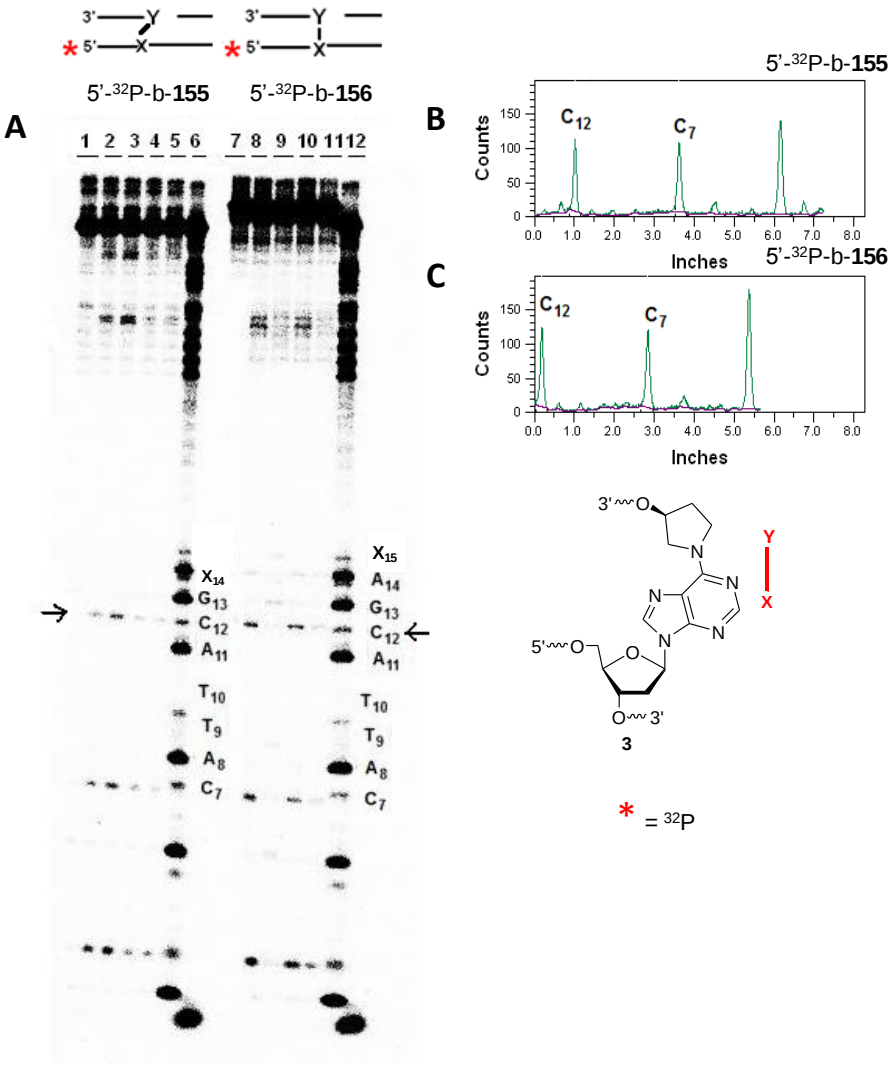
App. Figure 44. Hydroxyl radical cleavage reaction on 5'-^{32}P-b-**155** (observed ICL), 5'-^{32}P-b-**156** (unobserved ICL), and 5'-^{32}P-b-**157** (control duplex). (A) Hydroxyl radical cleavage reaction on 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156**, and 5'-^{32}P-b-**157** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 5'-^{32}P-b-**157** without any treatment. Lanes 2 – 5, 5'-^{32}P-b-**157** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lane 6, A-G sequencing of 5'-^{32}P-b-**157**. Lane 7, 5'-^{32}P-b-**155** without any treatment. Lanes 8 – 11, 5'-^{32}P-b-**155** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lane 12, A-G sequencing of 5'-^{32}P-b-**155**. Lane 13, 5'-^{32}P-b-**156** without any treatment. Lanes 14 – 17, 5'-^{32}P-b-**156** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lane 18, A-G sequencing of 5'-^{32}P-b-**156**. (B) Representative lane profile from hydroxyl radical cleavage of 5'-^{32}P-b-**157**. (C) Representative lane profile from hydroxyl radical cleavage of 5'-^{32}P-b-**155**. (D) Representative lane profile from hydroxyl radical cleavage of 5'-^{32}P-b-**156**.



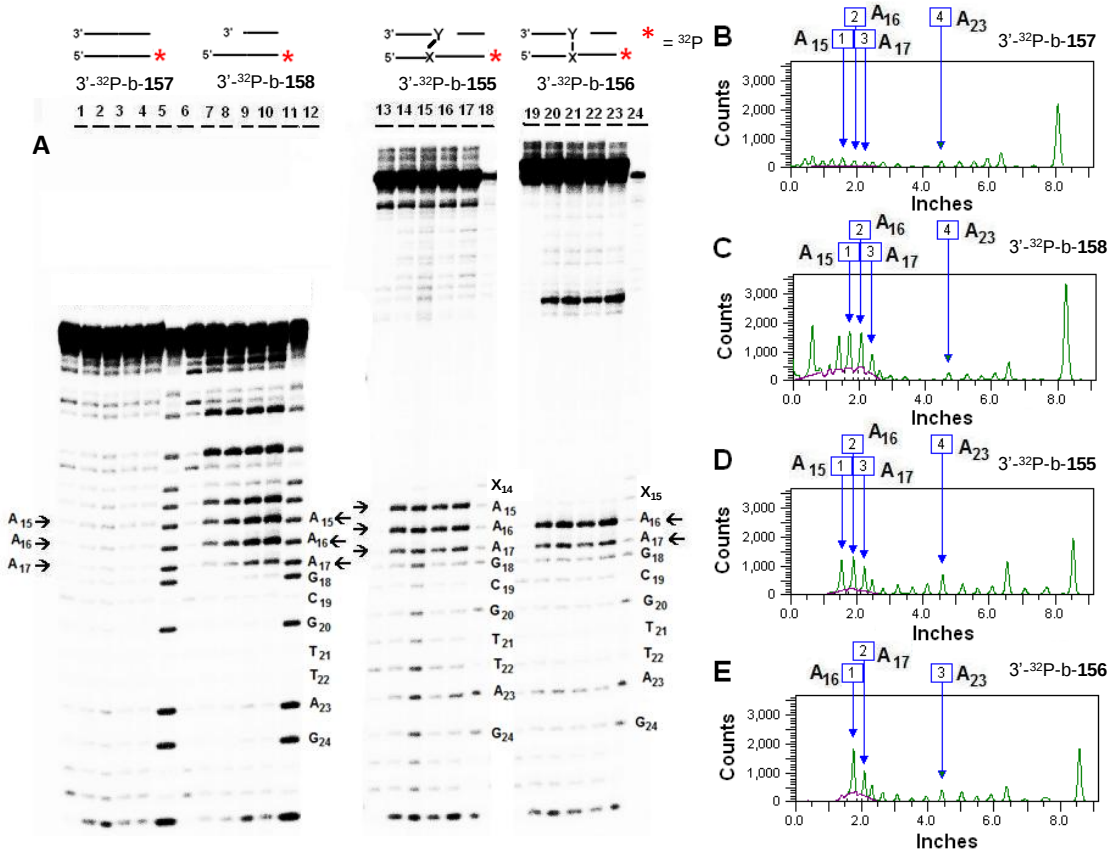
App. Figure 45. Diethylpyrocarbonate (DEPC) reaction on 5'-^{32}P-b-**155** (observed ICL), 5'-^{32}P-b-**156** (unobserved ICL), and 5'-^{32}P-b-**157** (control duplex). (A) Diethylpyrocarbonate (DEPC) reaction on 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156**, and 5'-^{32}P-b-**157** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 5'-^{32}P-b-**157**. Lanes 2 – 4, 5'-^{32}P-b-**157** treated with 1.04 M DEPC for 12 min, room temperature. Lane 5, A-G sequencing of 5'-^{32}P-b-**157**. Lane 6, 1 M piperidine treatment of 5'-^{32}P-b-**155**. Lanes 7 – 9, 5'-^{32}P-b-**155** treated with 1.04 M DEPC for 12 min, room temperature. Lane 10, A-G sequencing of 5'-^{32}P-b-**155**. Lane 11, 1 M piperidine treatment of 5'-^{32}P-b-**156**. Lanes 12 – 14, 5'-^{32}P-b-**156** treated with 1.04 M DEPC for 12 min, room temperature. Lane 15, A-G sequencing of 5'-^{32}P-b-**156**. (B) Representative lane profile from DEPC treatment of 5'-^{32}P-b-**157**. (C) Representative lane profile from DEPC treatment of 5'-^{32}P-b-**155**. (D) Representative lane profile from DEPC treatment of 5'-^{32}P-b-**156**.



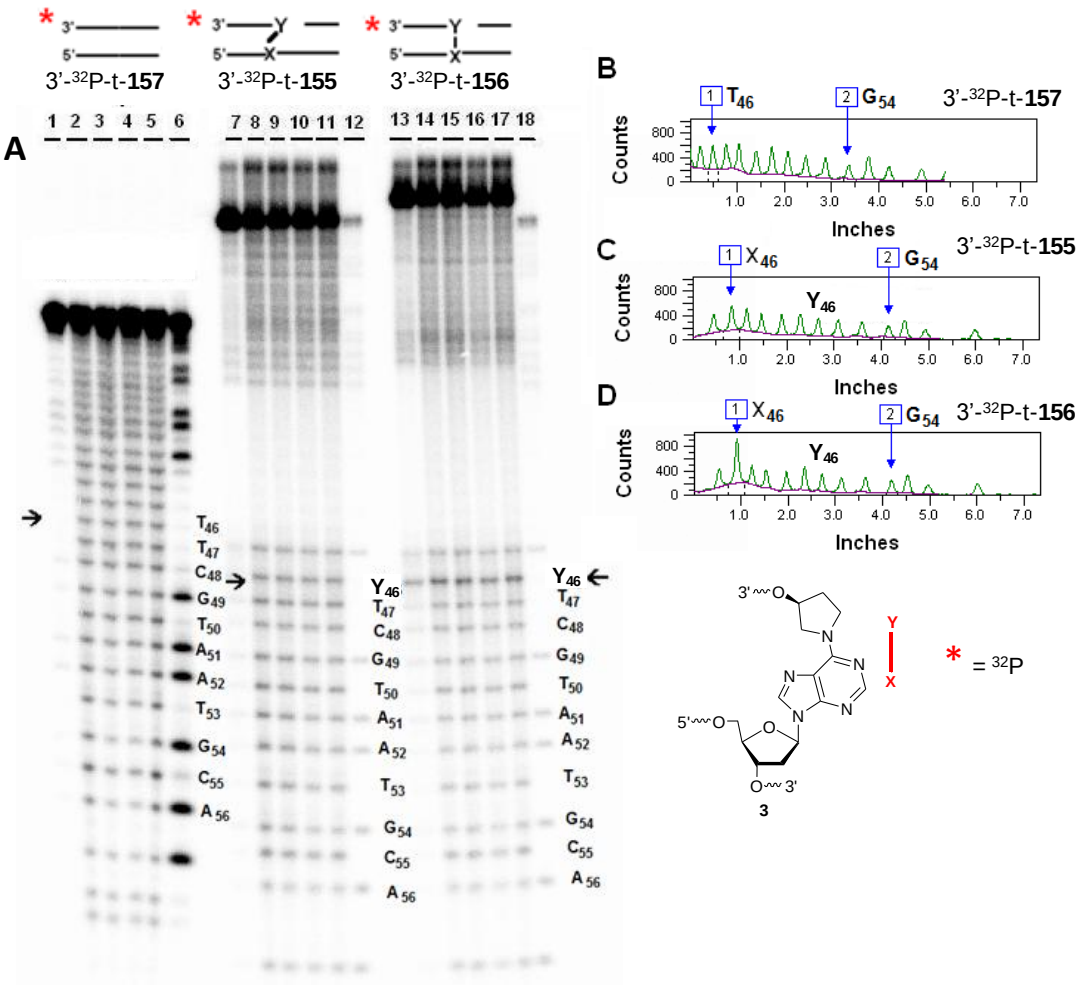
App. Figure 46. Dimethylsulfate (DMS) reaction on 5'-^{32}P-b-**155** (observed ICL), 5'-^{32}P-b-**156** (unobserved ICL), and 5'-^{32}P-b-**157** (control duplex). (A) DMS reaction on 5'-^{32}P-b-**155**, 5'-^{32}P-b-**156**, and 5'-^{32}P-b-**157** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 5'-^{32}P-b-**157**. Lanes 2 – 4, 5'-^{32}P-b-**157** treated with 25 mM DMS for 12 min, room temperature. Lane 5, A-G sequencing of 5'-^{32}P-b-**157**. Lane 6, 1 M piperidine treatment of 5'-^{32}P-b-**155**. Lanes 7 – 9, 5'-^{32}P-b-**155** treated with 25 mM DMS for 12 min, room temperature. Lane 10, A-G sequencing of 5'-^{32}P-b-**155**. Lane 11, 1 M piperidine treatment of 5'-^{32}P-b-**156**. Lanes 12 – 14, 5'-^{32}P-b-**156** treated with 25 mM DMS for 12 min, room temperature. Lane 15, A-G sequencing of 5'-^{32}P-b-**156**. (B) Representative lane profile from DMS treatment of 5'-^{32}P-b-**157**. (C) Representative lane profile from DMS treatment of 5'-^{32}P-b-**155**. (D) Representative lane profile from DMS treatment of 5'-^{32}P-b-**156**.



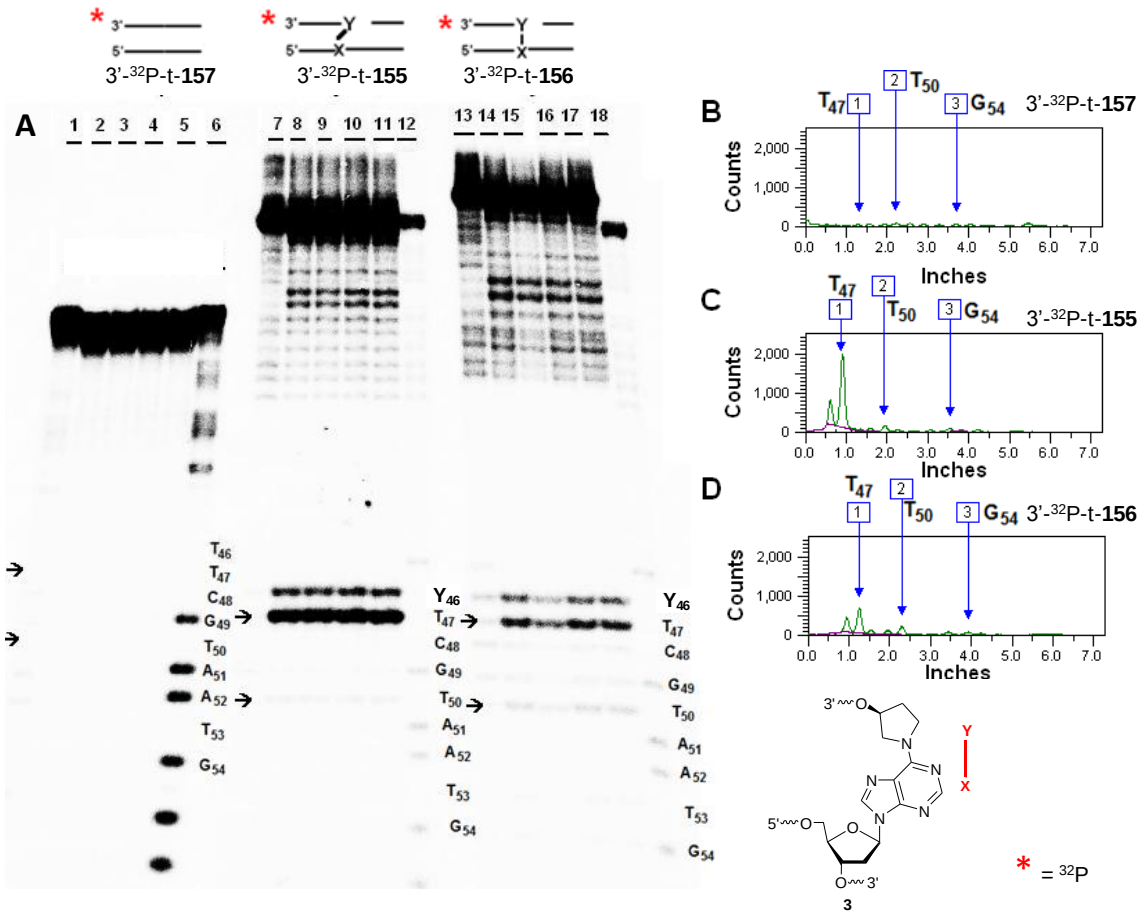
App. Figure 47. Hydroxylamine reaction on 5'-^{32}P-b-**155** (observed ICL), and 5'-^{32}P-b-**156** (unobserved ICL). (A) Hydroxylamine reaction on 5'-^{32}P-b-**155**, and 5'-^{32}P-b-**156** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 5'-^{32}P-b-**155**. Lanes 2 – 5, 5'-^{32}P-b-**155** treated with 0.80 M hydroxylamine for 10 min, room temperature. Lane 6, A-G sequencing of 5'-^{32}P-b-**155**. Lane 7, 1 M piperidine treatment of 5'-^{32}P-b-**156**. Lanes 8 – 11, 5'-^{32}P-b-**156** treated with 0.80 M hydroxylamine for 10 min, room temperature. Lane 12, A-G sequencing of 5'-^{32}P-b-**156**. (B) Representative lane profile from hydroxylamine treatment of 5'-^{32}P-b-**155**. (C) Representative lane profile from hydroxylamine treatment of 5'-^{32}P-b-**156**.



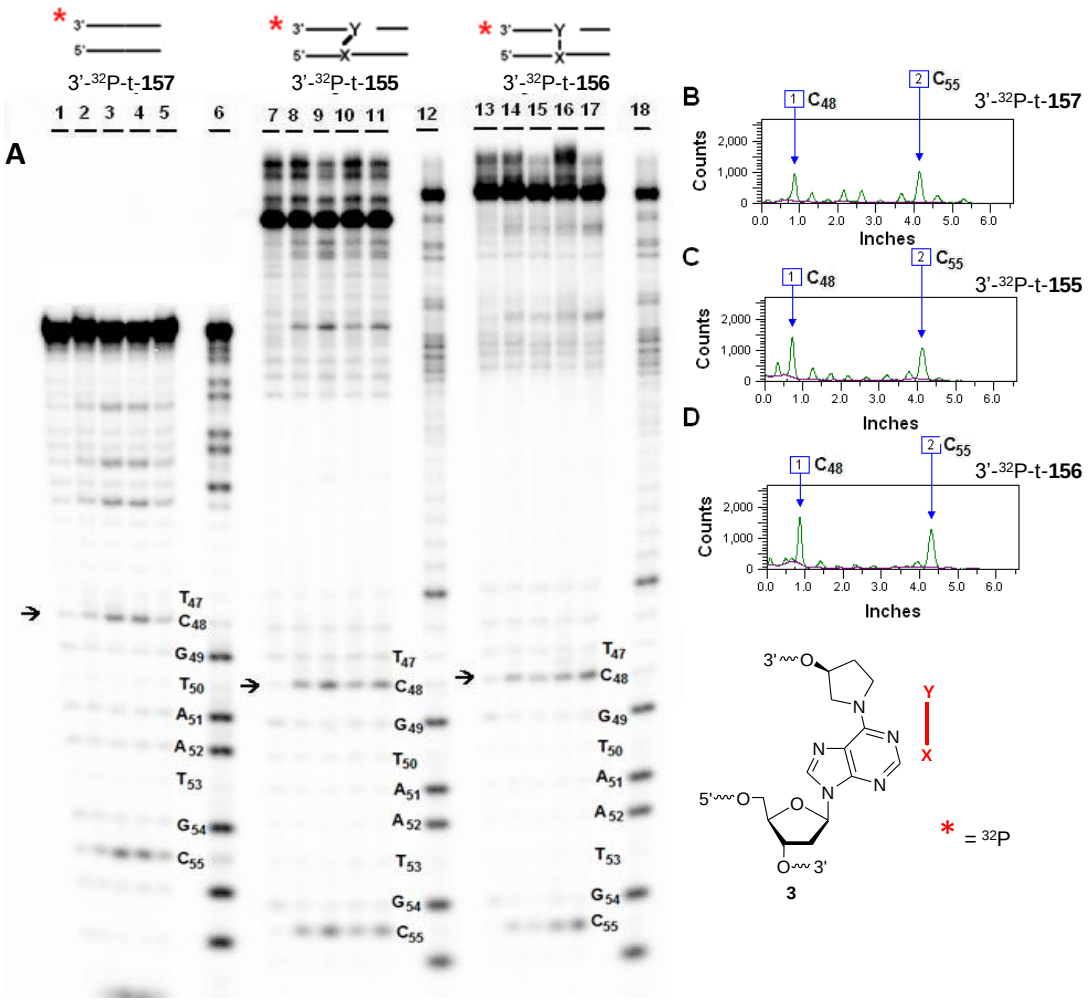
App. Figure 48. Diethylpyrocarbonate (DEPC) reaction on 3'-^{32}P-b-**155** (observed ICL), 3'-^{32}P-b-**156** (unobserved ICL), 3'-^{32}P-b-**157** (control duplex), and 3'-^{32}P-b-**158** (control). (A) Diethylpyrocarbonate (DEPC) reaction on 3'-^{32}P-b-**155**, 3'-^{32}P-b-**156**, 3'-^{32}P-b-**157** and 3'-^{32}P-b-**158**. analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 3'-^{32}P-b-**157**. Lanes 2 – 5, 3'-^{32}P-b-**157** treated with 1.04 M DEPC for 12 min, room temperature. Lane 6, A-G sequencing of 3'-^{32}P-b-**157**. Lane 7, 1 M piperidine treatment of 3'-^{32}P-b-**158**. Lanes 8 – 11, 3'-^{32}P-b-**158** treated with 1.04 M DEPC for 12 min, room temperature. Lane 12, A-G sequencing of 5'-^{32}P-b-**158**. Lane 13, 1 M piperidine treatment of 3'-^{32}P-b-**155**. Lanes 14 – 17, 3'-^{32}P-b-**155** treated with 1.04 M DEPC for 12 min, room temperature. Lane 18, A-G sequencing of 5'-^{32}P-b-**155**. Lane 19, 1 M piperidine treatment of 3'-^{32}P-b-**156**. Lanes 20 – 23, 3'-^{32}P-b-**156** treated with 1.04 M DEPC for 12 min, room temperature. Lane 24, A-G sequencing of 3'-^{32}P-b-**156**. (B) Representative lane profile from DEPC treatment of 3'-^{32}P-b-**157**. (C) Representative lane profile from DEPC treatment of 3'-^{32}P-b-**158**. (D) Representative lane profile from DEPC treatment of 3'-^{32}P-b-**155**. (E) Representative lane profile from DEPC treatment of 3'-^{32}P-b-**156**.



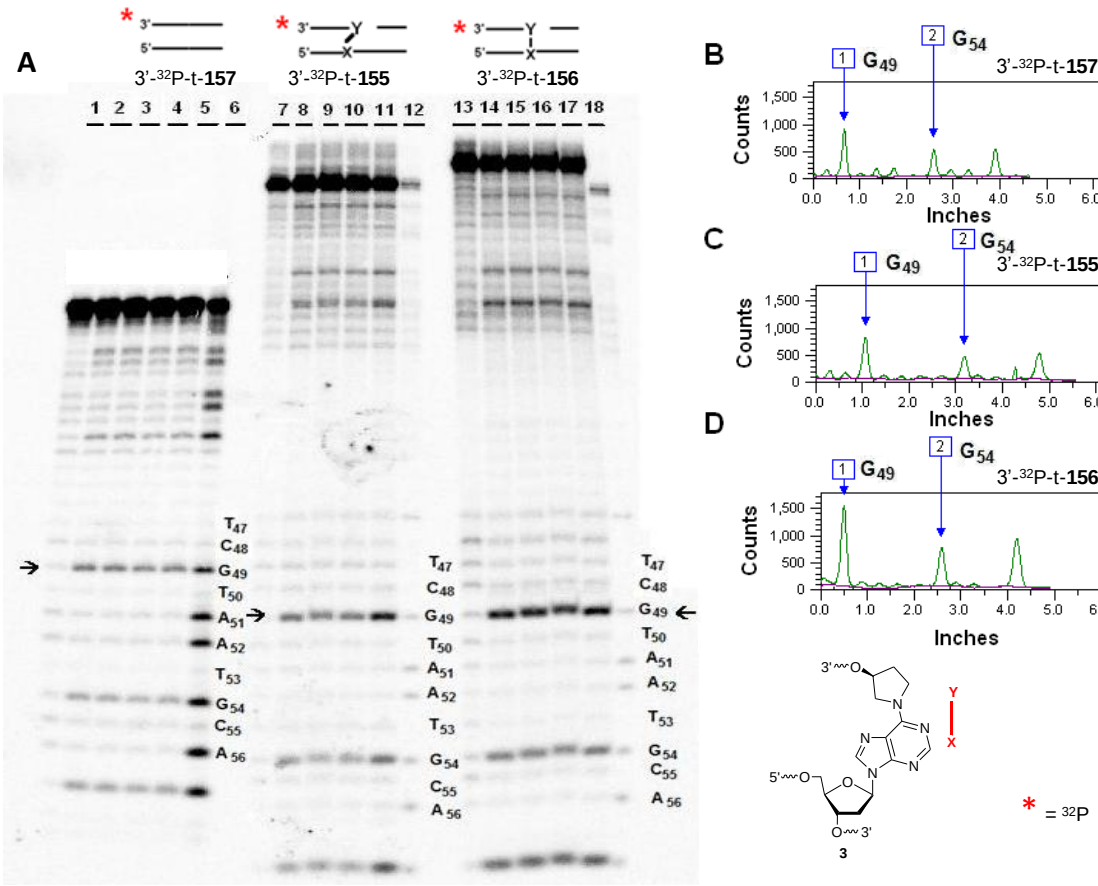
App. Figure 49. Hydroxyl radical cleavage reaction on 3'-^{32}P-t-**155** (observed ICL), 3'-^{32}P-t-**156** (unobserved ICL), and 3'-^{32}P-t-**157** (control duplex). (A) Hydroxyl radical cleavage reaction on 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, and 3'-^{32}P-t-**157** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 3'-^{32}P-t-**157** without any treatment. Lanes 2 – 5, 3'-^{32}P-t-**157** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lane 6, A-G sequencing of 3'-^{32}P-t-**157**. Lane 7, 3'-^{32}P-t-**155** without any treatment. Lanes 8 – 11, 3'-^{32}P-t-**155** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lane 12, A-G sequencing of 3'-^{32}P-t-**155**. Lane 13, 3'-^{32}P-t-**156** without any treatment. Lanes 14 – 17, 3'-^{32}P-t-**156** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lane 18, A-G sequencing of 3'-^{32}P-t-**156**. (B) Representative lane profile from hydroxyl radical cleavage of 3'-^{32}P-t-**157**. (C) Representative lane profile from hydroxyl radical cleavage of 3'-^{32}P-t-**155**. (D) Representative lane profile from hydroxyl radical cleavage of 3'-^{32}P-t-**156**.



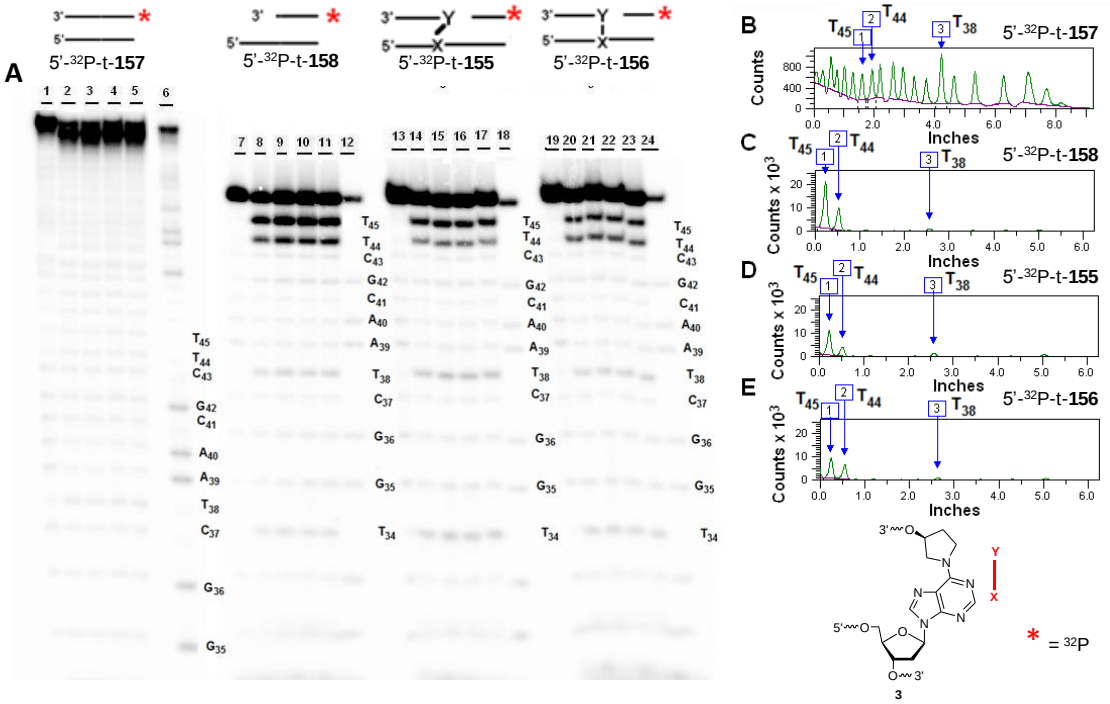
App. Figure 50. $KMnO_4$ cleavage reaction on 3'-^{32}P-t-**155** (observed ICL), 3'-^{32}P-t-**156** (unobserved ICL), and 3'-^{32}P-t-**157** (control duplex). (A) $KMnO_4$ reaction on 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, and 3'-^{32}P-t-**157** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 3'-^{32}P-t-**157**. Lanes 2 – 5, 3'-^{32}P-t-**157** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 6, A-G sequencing of 3'-^{32}P-t-**157**. Lane 7, 1 M piperidine treatment of 3'-^{32}P-t-**155**. Lanes 8 – 11, 3'-^{32}P-t-**155** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 12, A-G sequencing of 3'-^{32}P-t-**155**. Lane 13, 1 M piperidine treatment of 3'-^{32}P-t-**156**. Lanes 14 – 17, 3'-^{32}P-t-**156** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 18, A-G sequencing of 3'-^{32}P-t-**156**. (B) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**157**. (C) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**155**. (D) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**156**.



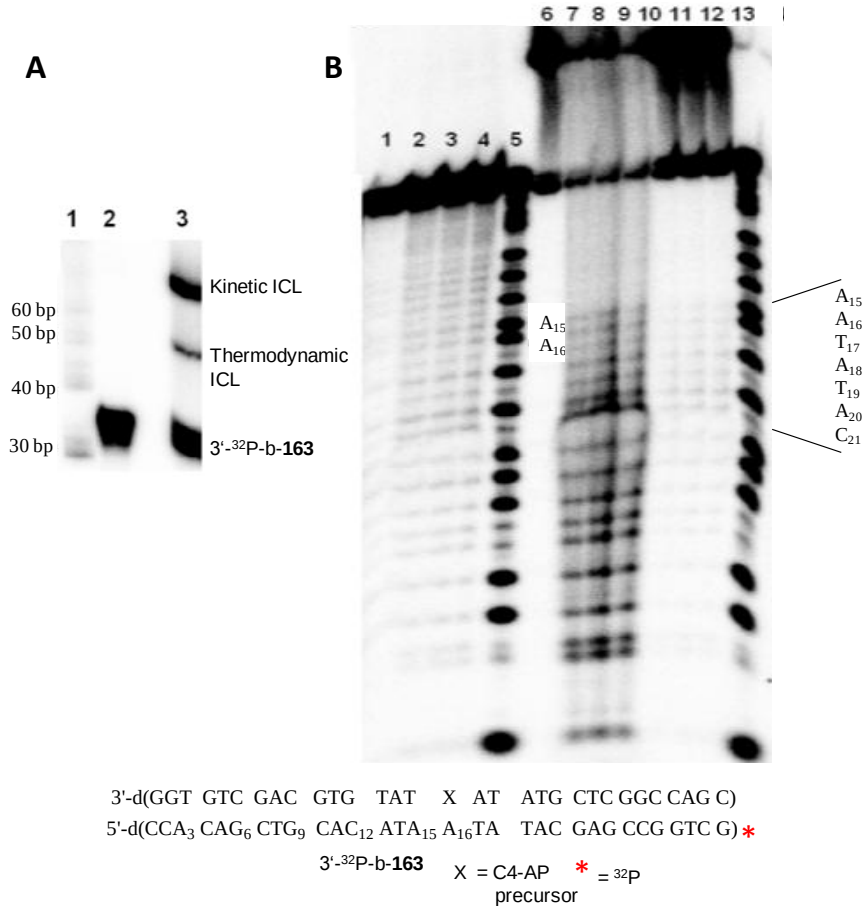
App. Figure 51. Hydroxylamine cleavage reaction on 3'-^{32}P-t-**155** (observed ICL), 3'-^{32}P-t-**156** (unobserved ICL), and 3'-^{32}P-t-**157** (control duplex). (A) Hydroxylamine reaction on 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, and 3'-^{32}P-t-**157** analyzed on 20% denaturing PAGE. Nucleotide of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 3'-^{32}P-t-**157**. Lanes 2 – 5, 3'-^{32}P-t-**157** subjected to 0.80 M hydroxylamine for 10 min, room temperature. Lane 6, A-G sequencing of 3'-^{32}P-t-**157**. Lane 7, 1 M piperidine treatment of 3'-^{32}P-t-**155**. Lanes 8 – 11, 3'-^{32}P-t-**155** subjected to 0.80 M hydroxylamine for 10 min, room temperature. Lane 12, A-G sequencing of 3'-^{32}P-t-**155**. Lane 13, 1 M piperidine treatment of 3'-^{32}P-t-**156**. Lanes 14 – 17, 3'-^{32}P-t-**156** subjected to 0.80 M hydroxylamine for 10 min, room temperature. Lane 18, A-G sequencing of 3'-^{32}P-t-**156**. (B) Representative lane profile from hydroxylamine reaction of 3'-^{32}P-t-**157**. (C) Representative lane profile from hydroxylamine reaction of 3'-^{32}P-t-**155**. (D) Representative lane profile from hydroxylamine reaction of 3'-^{32}P-t-**156**.



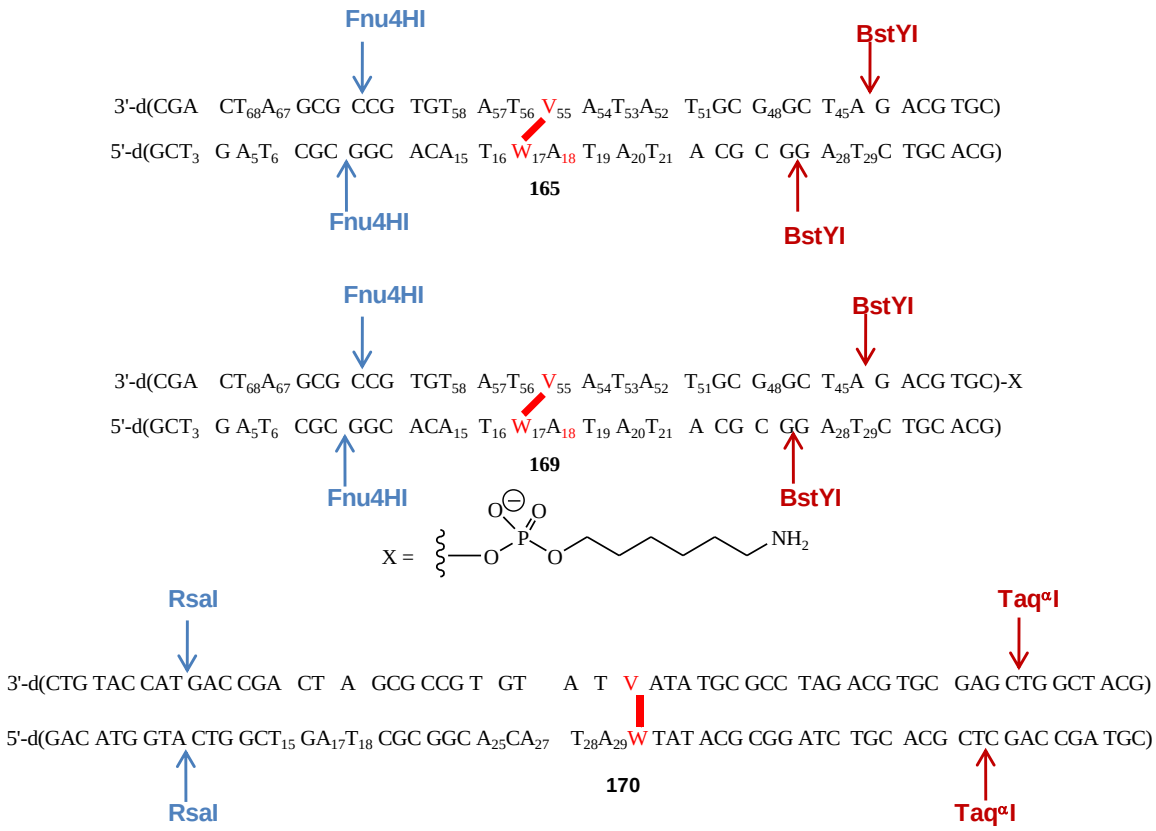
App. Figure 52. DMS cleavage reaction on 3'-^{32}P-t-**155** (observed ICL), 3'-^{32}P-t-**156** (unobserved ICL), and 3'-^{32}P-t-**157** (control duplex). (A) DMS reaction on 3'-^{32}P-t-**155**, 3'-^{32}P-t-**156**, and 3'-^{32}P-t-**157** analyzed on 20% denaturing PAGE. Nucleotides of interest are indicated with arrows. Lane 1, 1 M piperidine treatment of 3'-^{32}P-t-**157**. Lanes 2 – 5, 3'-^{32}P-t-**157** subjected to 25 mM DMS for 12 min, room temperature. Lane 6, A-G sequencing of 3'-^{32}P-t-**157**. Lane 7, 1 M piperidine treatment of 3'-^{32}P-t-**155**. Lanes 8 – 11, 3'-^{32}P-t-**155** subjected to 25 mM DMS for 12 min, room temperature. Lane 12, A-G sequencing of 3'-^{32}P-t-**155**. Lane 13, 1 M piperidine treatment of 3'-^{32}P-t-**156**. Lanes 14 – 17, 3'-^{32}P-t-**156** subjected to 25 mM DMS for 12 min, room temperature. Lane 18, A-G sequencing of 3'-^{32}P-t-**156**. (B) Representative lane profile from DMS reaction of 3'-^{32}P-t-**157**. (C) Representative lane profile from DMS reaction of 3'-^{32}P-t-**155**. (D) Representative lane profile from DMS reaction of 3'-^{32}P-t-**156**.



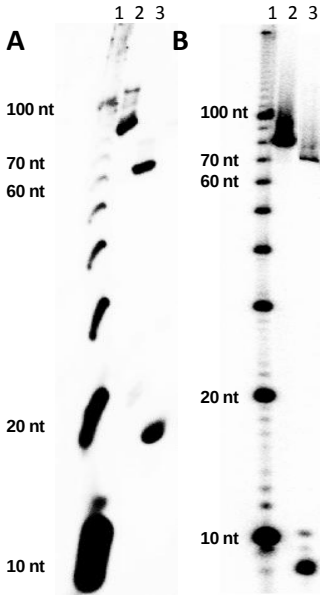
App. Figure 53. $KMnO_4$ cleavage reaction on 5'-^{32}P-t-**155** (observed ICL), 5'-^{32}P-t-**156** (unobserved ICL), 5'-^{32}P-t-**157** (control duplex) and 5'-^{32}P-t-**158** (control). (A) $KMnO_4$ reaction on 5'-^{32}P-t-**155**, 5'-^{32}P-t-**156**, 5'-^{32}P-t-**157** and 5'-^{32}P-t-**158** analyzed on 20% denaturing PAGE. Lane 1, 1 M piperidine treatment of 5'-^{32}P-t-**157**. Lanes 2 – 5, 5'-^{32}P-t-**157** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 6, A-G sequencing of 5'-^{32}P-t-**157**. Lane 7, 1 M piperidine treatment of 5'-^{32}P-t-**158**. Lanes 8 – 11, 5'-^{32}P-t-**158** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 12, A-G sequencing of 5'-^{32}P-t-**158**. Lane 13, 1 M piperidine treatment of 5'-^{32}P-t-**155**. Lanes 14 – 17, 5'-^{32}P-t-**155** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 18, A-G sequencing of 5'-^{32}P-t-**155**. Lane 19, 1 M piperidine treatment of 5'-^{32}P-t-**156**. Lanes 20 – 23, 5'-^{32}P-t-**156** subjected to 0.3 mM $KMnO_4$ for 10 min, room temperature. Lane 24, A-G sequencing of 5'-^{32}P-t-**156**. (B) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**157**. (C) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**158**. (D) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**155**. (E) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**156**.



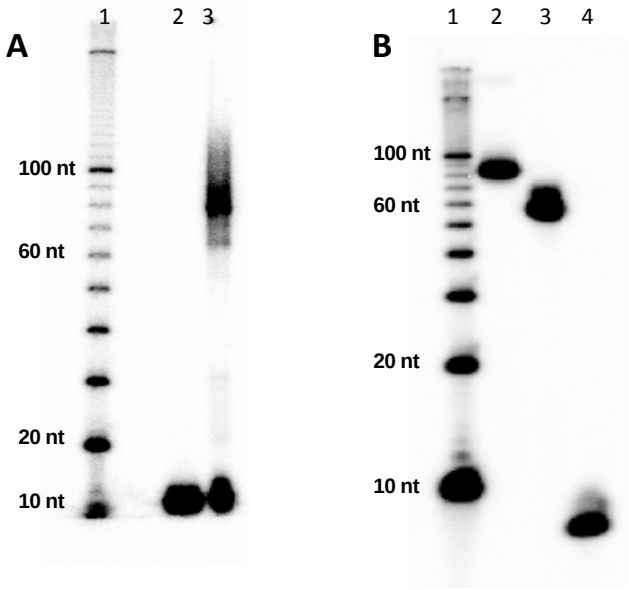
App. Figure 54. Photoirradiation (350 nm) of 3'-^{32}P-b-**163** and hydroxyl radical cleavage reaction on C4-AP kinetic ICL. (A) Photoirradiation (350 nm) of 3'-^{32}P-b-**163**. Positions of kinetic (high molecular weight) and thermodynamic (low molecular weight) C4-AP ICLs are annotated. Lane 1, 10 bp ladder. Lane 2, 3'-^{32}P-b-**163** without any treatment. Lane 3, 3'-^{32}P-b-**163** subjected to 45 min irradiation (350 nm) followed by overnight incubation (37 °C). (B) Hydroxyl radical cleavage of C4-AP kinetic ICL. Lane 1, 3'-^{32}P-b-**163** without any treatment. Lane 2, 3'-^{32}P-b-**163** without any treatment. Lanes 2 – 4, 3'-^{32}P-b-**163** subjected to hydroxyl radical cleavage, 5 min, room temperature. Lane 5, A-G sequencing of 3'-^{32}P-b-**163**. Lane 6, the slowest moving band isolated from irradiation of 3'-^{32}P-b-**163** without any treatment. Lanes 7 – 9, hydroxyl radical cleavage treatment of slowest moving band from irradiation of 3'-^{32}P-b-**163**, 8 min, room temperature. Lanes 10 – 12, hydroxyl radical cleavage treatment of slowest moving band from irradiation of 3'-^{32}P-b-**163**, 5 min, room temperature. Lane 13, A-G sequencing of 3'-^{32}P-b-**163**.



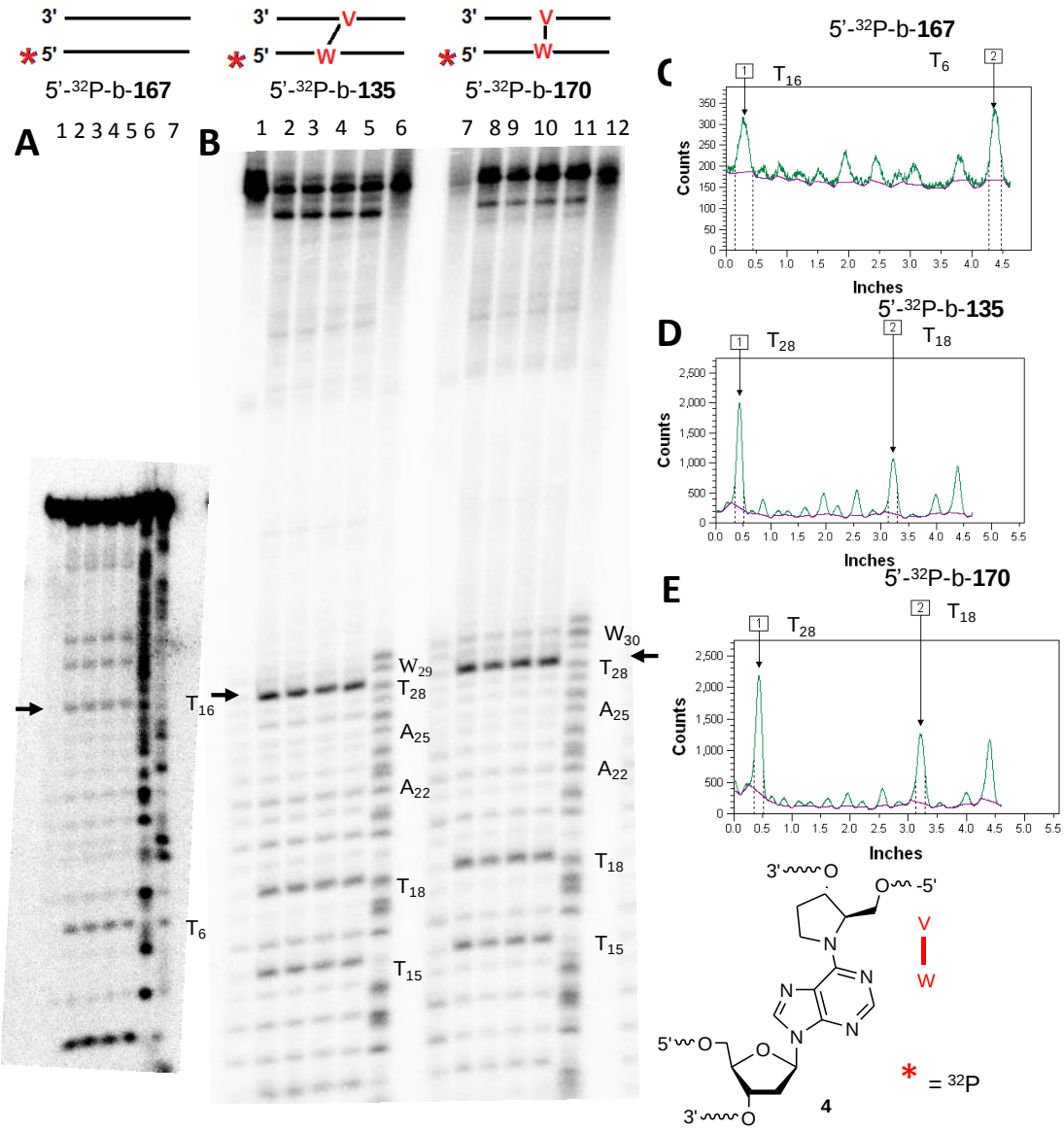
App. Figure 55. Restriction enzyme cleavage sites on **165, 169** and **170.**



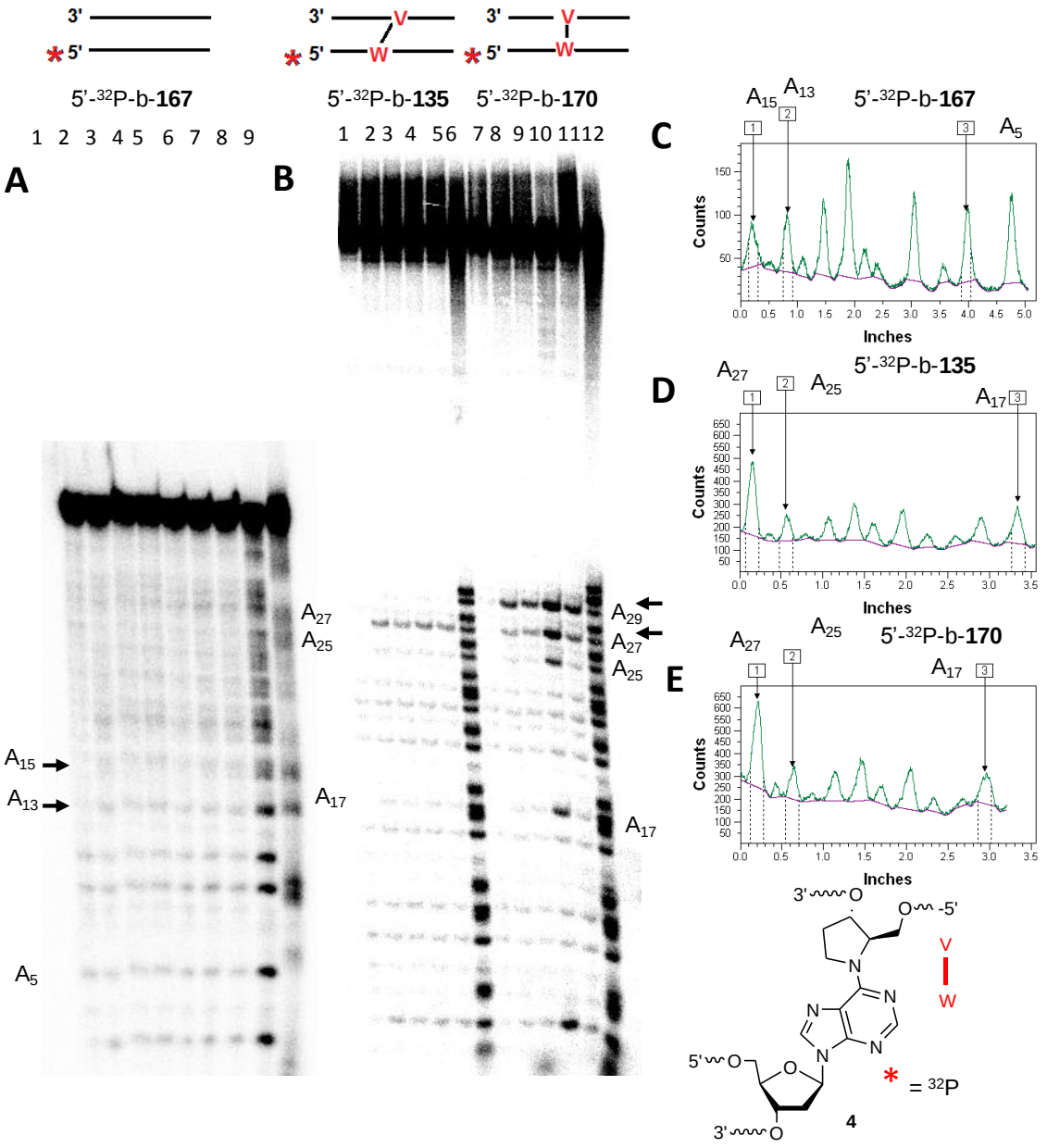
App. Figure 56. Enzyme treatment of purified 5'-^{32}P-b-**165** and 5'-^{32}P-b-**169**. (A) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of purified 5'-^{32}P-b-**165.** Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-b-**165** without any treatment. Lane 3, 5'-^{32}P-b-**165** treated with Fnu4HI. (B) Representative 20% denaturing PAGE gel analysis of restriction enzyme treatment of purified 5'-^{32}P-b-**169** Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-b-**169** without any treatment. Lane 3, 5'-^{32}P-b-**169** treated with Fnu4HI. Lane 4, 5'-^{32}P-b-**169** treated with BstYI.



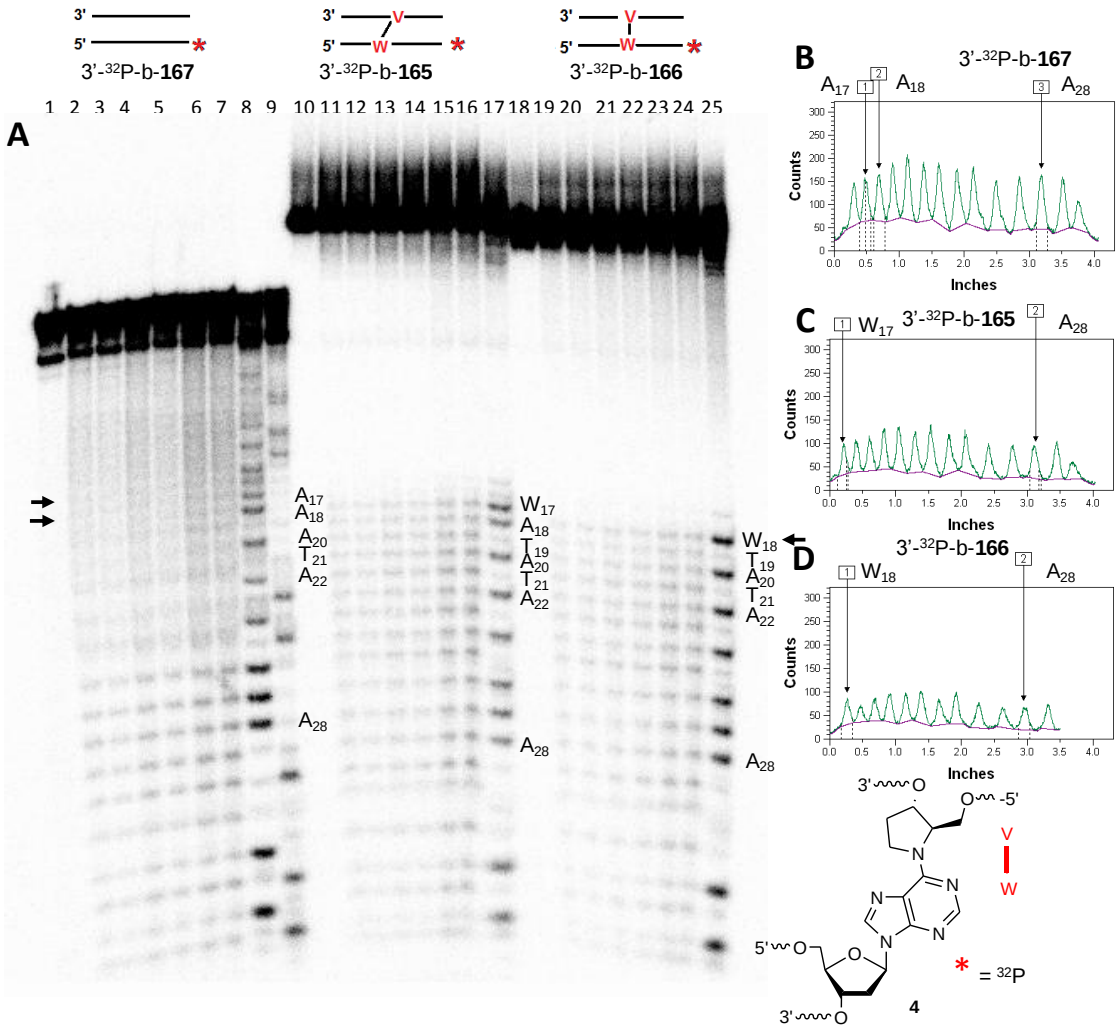
App. Figure 57. Ligation for constructing 5'-^{32}P-b-**170** and restriction enzyme treatment of purified 5'-^{32}P-b-**170**. (A) Representative 15% denaturing PAGE gel showing the ligation to construct 5'-^{32}P-b-**170.** Lane 1, 10 base pair ladder. Lane 2**,** 5'-^{32}P-b-**131.** Lane 3, crude reaction to produce 5'-^{32}P-b-**170**. (B) Representative 15% denaturing PAGE gel analysis of restriction enzyme treatment of purified 5'-^{32}P-b-**170.** Lane 1, 10 base pair DNA ladder. Lane 2, 5'-^{32}P-b-**170** without any treatment. Lane 3, 5'-^{32}P-b-**170** treated with RsaI. Lane 4, , 5'-^{32}P-b-**170** treated with Taq$^{\alpha}$I.



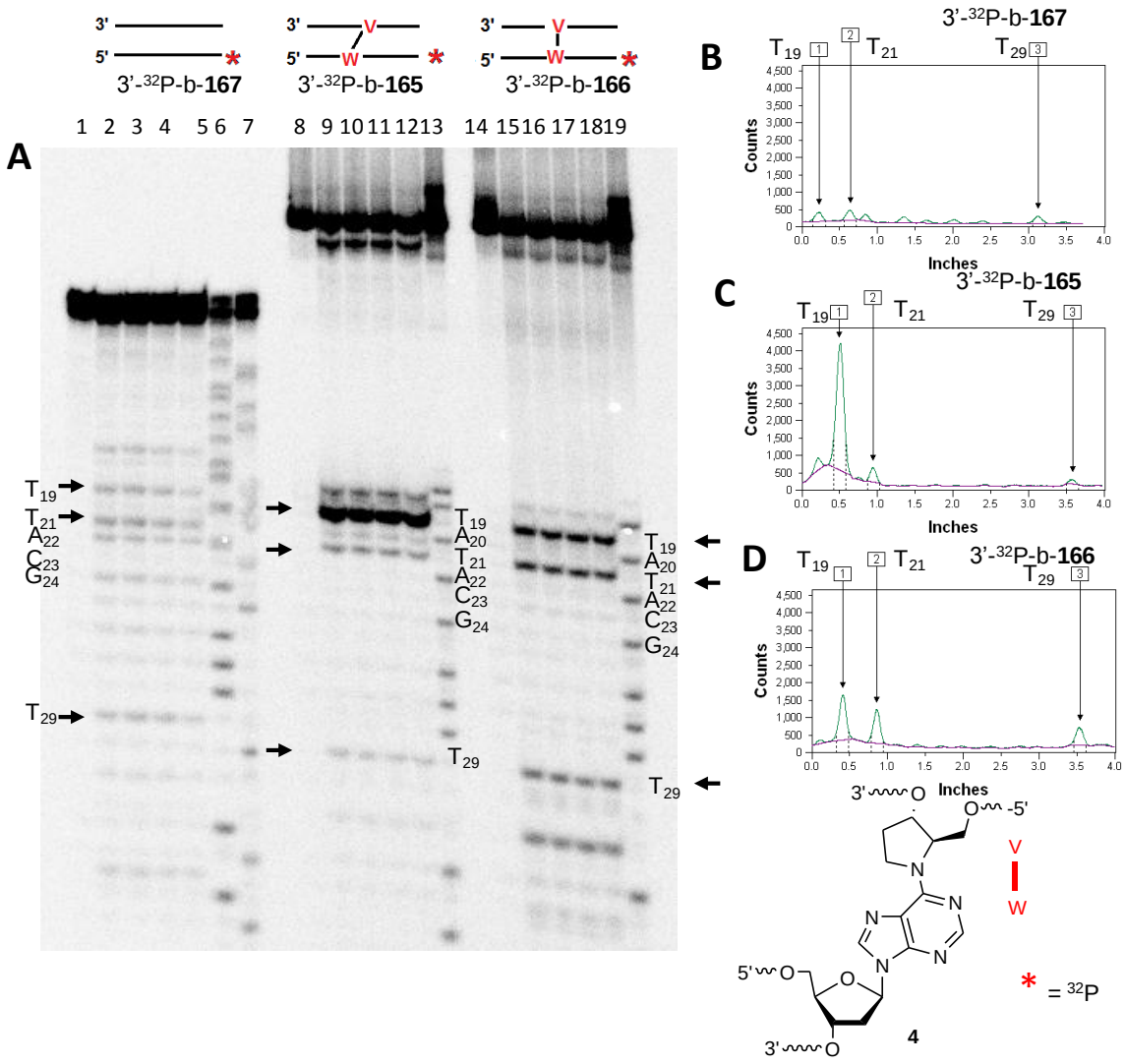
App. Figure 58. $KMnO_4$ reaction on 5'-^{32}P-b-**167** (control duplex), 5'-^{32}P-b-**135** (observed ICL), and 5'-^{32}P-b-**170** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) $KMnO_4$ reaction on 3'-^{32}P-b-**167** analyzed on 20% denaturing PAGE. Lane 1, 1 M piperidine treatment of 5'-^{32}P-b-**167**. Lanes 2 – 5, 5'-^{32}P-b-**167** treated with 0.6 mM $KMnO_4$ for 10 min, room temperature. Lanes 6 and 7, A-G and T-C sequencing of 5'-^{32}P-b-**167**, respectively. (B) $KMnO_4$ reaction on 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170** analyzed on 20% denaturing PAGE. Lanes 1 and 7, 1 M piperidine treatment of 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170**, respectively. Lanes 2 – 5 and 8 – 11, 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170** treated with 0.6 mM $KMnO_4$ for 10 min (room temperature), respectively. Lanes 6 and 12, A-G sequencing of 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170**, respectively. (C) Representative lane profile from $KMnO_4$ treatment of 5'-^{32}P-b-**167**. (D) Representative lane profile from $KMnO_4$ treatment of 5'-^{32}P-b-**135**. (E) Representative lane profile from $KMnO_4$ treatment of 5'-^{32}P-b-**170**.



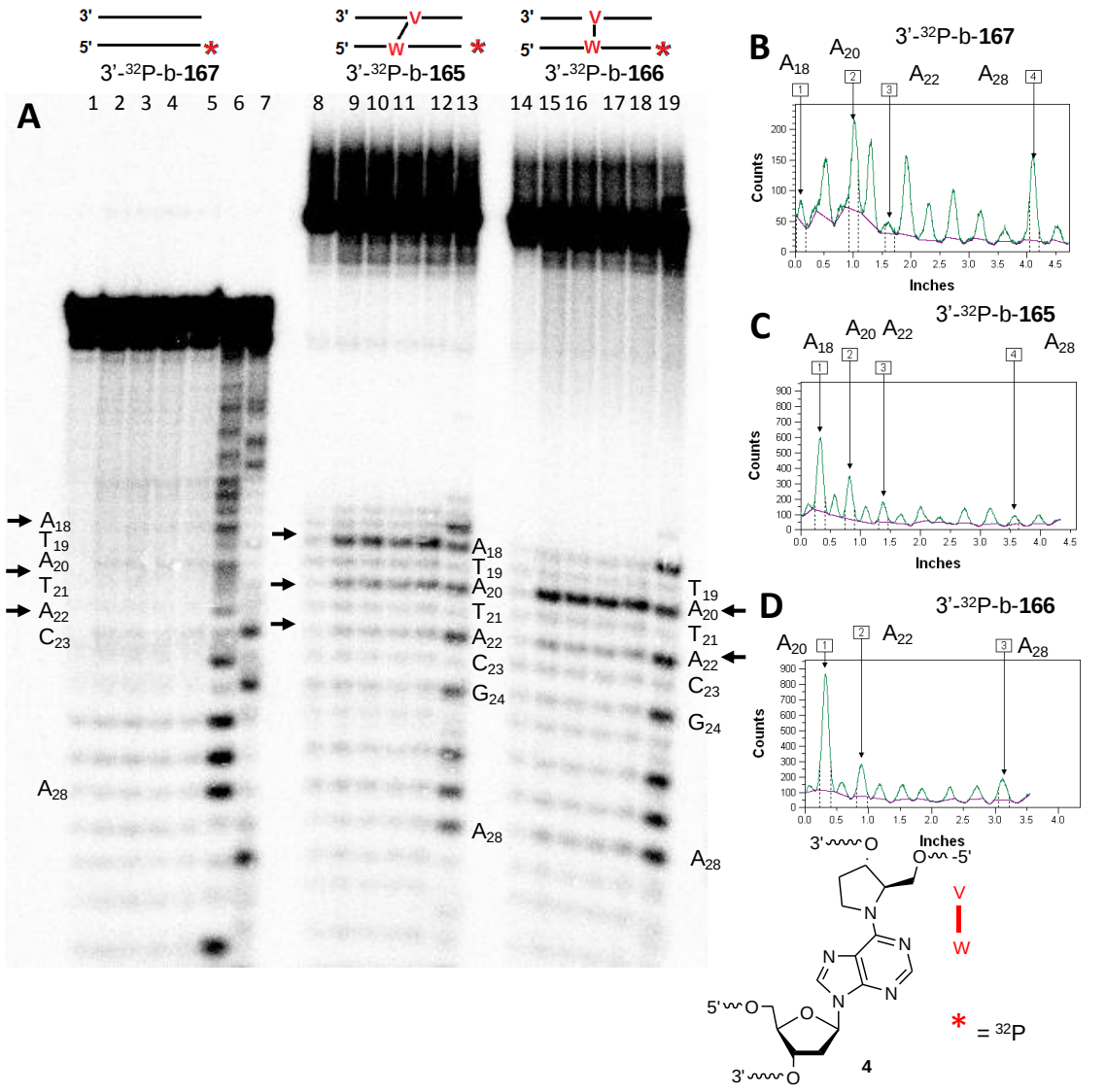
App. Figure 59. Diethylpyrocarbonate (DEPC) reaction on 5'-^{32}P-b-**167** (control duplex), 5'-^{32}P-b-**135** (observed ICL), and 5'-^{32}P-b-**170** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) DEPC reaction on 3'-^{32}P-b-**167** analyzed on 20% denaturing PAGE. Lane 1, 1 M piperidine treatment of 5'-^{32}P-b-**167**. Lanes 2 – 7, 5'-^{32}P-b-**167** treated with 1.75 M DEPC for 12 min, room temperature. Lanes 8 and 9, A-G and T-C sequencing of 5'-^{32}P-b-**167**, respectively. (B) Diethylpyrocarbonate (DEPC) reaction on 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170** analyzed on 20% denaturing PAGE. Lanes 1 and 7, 1 M piperidine treatment of 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170**, respectively. Lanes 2 – 5 and 8 – 11, 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170** treated with 1.75 M DEPC for 12 min (room temperature), respectively. Lanes 6 and 12, A-G sequencing of 5'-^{32}P-b-**135**, and 5'-^{32}P-b-**170**, respectively. (C) Representative lane profile from DEPC treatment of 5'-^{32}P-b-**167**. (D) Representative lane profile from DEPC treatment of 5'-^{32}P-b-**135**. (E) Representative lane profile from DEPC treatment of 5'-^{32}P-b-**170**.



App. Figure 60. Hydroxyl radical cleavage reaction on 3'-^{32}P-b-**167** (control duplex), 3'-^{32}P-b-**165** (observed ICL), and 3'-^{32}P-b-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) Hydroxyl radical cleavage reaction on 3'-^{32}P-b-**165**, 3'-^{32}P-b-**166**, and 3'-^{32}P-b-**167** analyzed on 20% denaturing PAGE. Lanes 1, 10 and 18, 3'-^{32}P-b-**167,** 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** without any treatment. Lanes 2 – 7, 11 – 16 and 19 – 24, 3'-^{32}P-b-**167,** 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lanes 8 and 9, A-G and T-C sequencing of 3'-^{32}P-b-**167**, respectively. Lanes 17 and 25, A-G sequencing of 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**, respectively. (B) Representative lane profile from hydroxyl radical cleavage reaction of 3'-^{32}P-b-**167**. (C) Representative lane profile from hydroxyl radical cleavage reaction of 3'-^{32}P-b-**165.** (D) Representative lane profile from hydroxyl radical cleavage reaction of 3'-^{32}P-b-**166**.

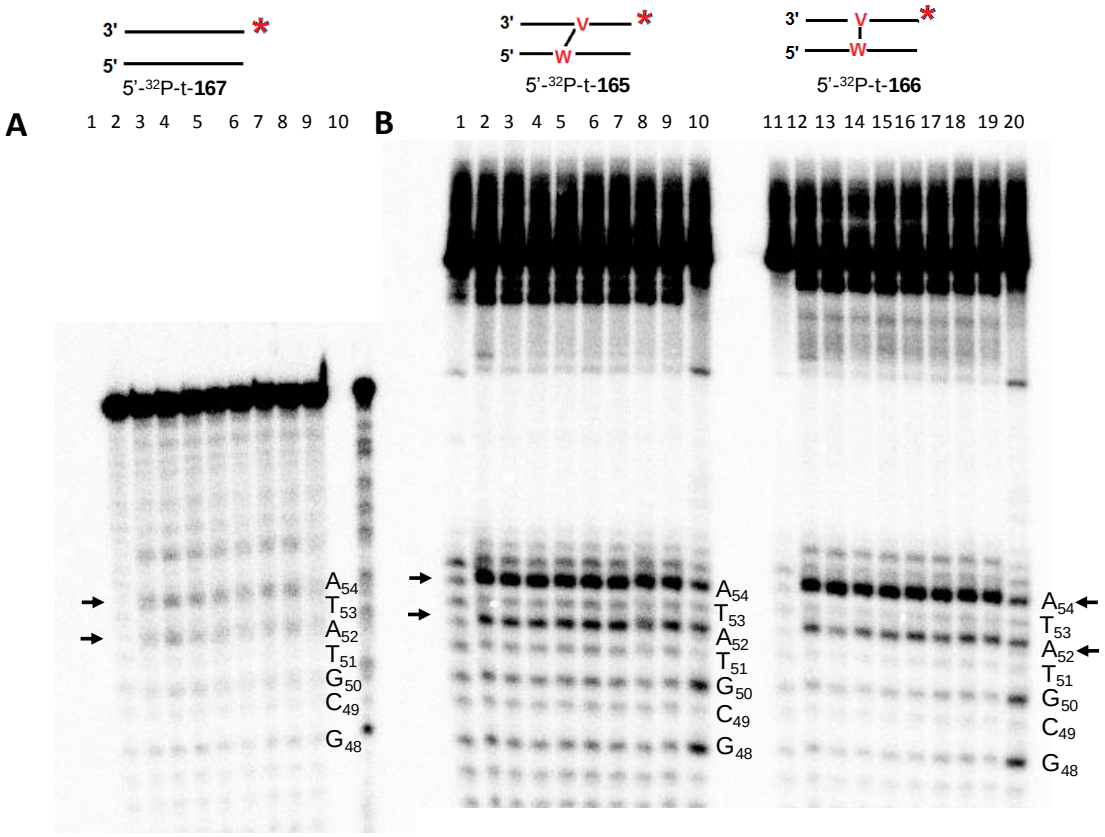


App. Figure 61. $KMnO_4$ reaction on 3'-^{32}P-b-**167** (control duplex), 3'-^{32}P-b-**165** (observed ICL), and 3'-^{32}P-b-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) $KMnO_4$ reaction on 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**, and 3'-^{32}P-b-**167** analyzed on 20% denaturing PAGE. Lanes 1, 8, and 14, 1 M piperidine treatment of 3'-^{32}P-b-**167,** 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**. Lanes 2 – 5, 9 – 12 and 15 – 18, 3'-^{32}P-b-**167,** 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** treated with 0.6 mM $KMnO_4$ for 10 min, room temperature. Lanes 6 and 7, A-G and T-C sequencing of 3'-^{32}P-b-**167**, respectively. Lanes 13 and 19, A-G sequencing of 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**, respectively. (B) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-b-**167**. (C) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-b-**165**. (D) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-b-**166.**

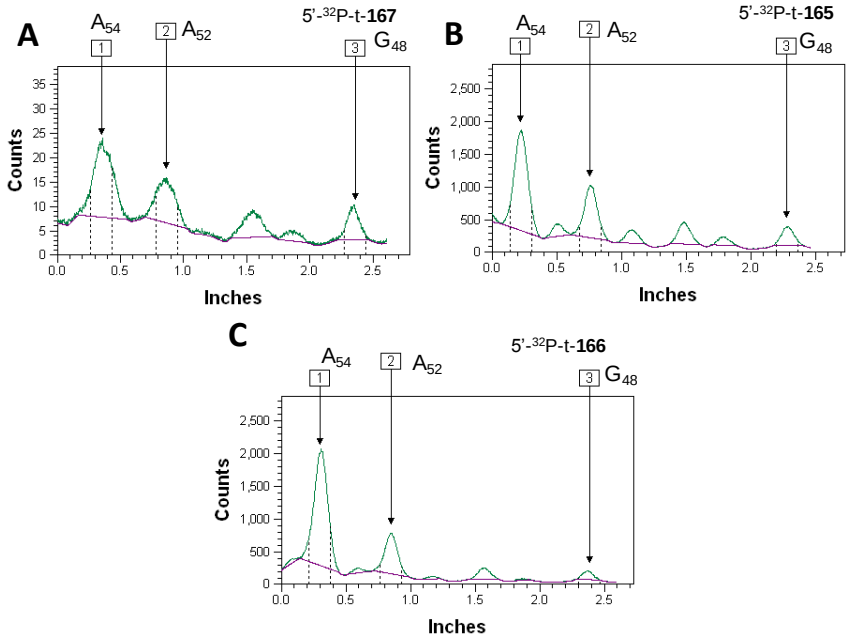


App. Figure 62. DEPC reaction on 3'-^{32}P-b-**167** (control duplex), 3'-^{32}P-b-**165** (observed ICL), and 3'-^{32}P-b-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) DEPC reaction on 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**, and 3'-^{32}P-b-**167** analyzed on 20% denaturing PAGE. Lanes 1, 8, and 14, 1 M piperidine treatment of 3'-^{32}P-b-**167,** 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**. Lanes 2 – 5, 9 – 12 and 15 – 18, 3'-^{32}P-b-**167,** 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166** treated with 1.75 M DEPC for 12 min, room temperature. Lanes 6 and 7, A-G and T-C sequencing of 3'-^{32}P-b-**167**, respectively. Lanes 13 and 19, A-G sequencing of 3'-^{32}P-b-**165**, and 3'-^{32}P-b-**166**, respectively. (C) Representative lane profile from DEPC reaction of 3'-^{32}P-b-**167**. (D) Representative lane profile from DEPC reaction of 3'-^{32}P-b-**165**. (E) Representative lane profile from DEPC reaction of 3'-^{32}P-b-**166**.

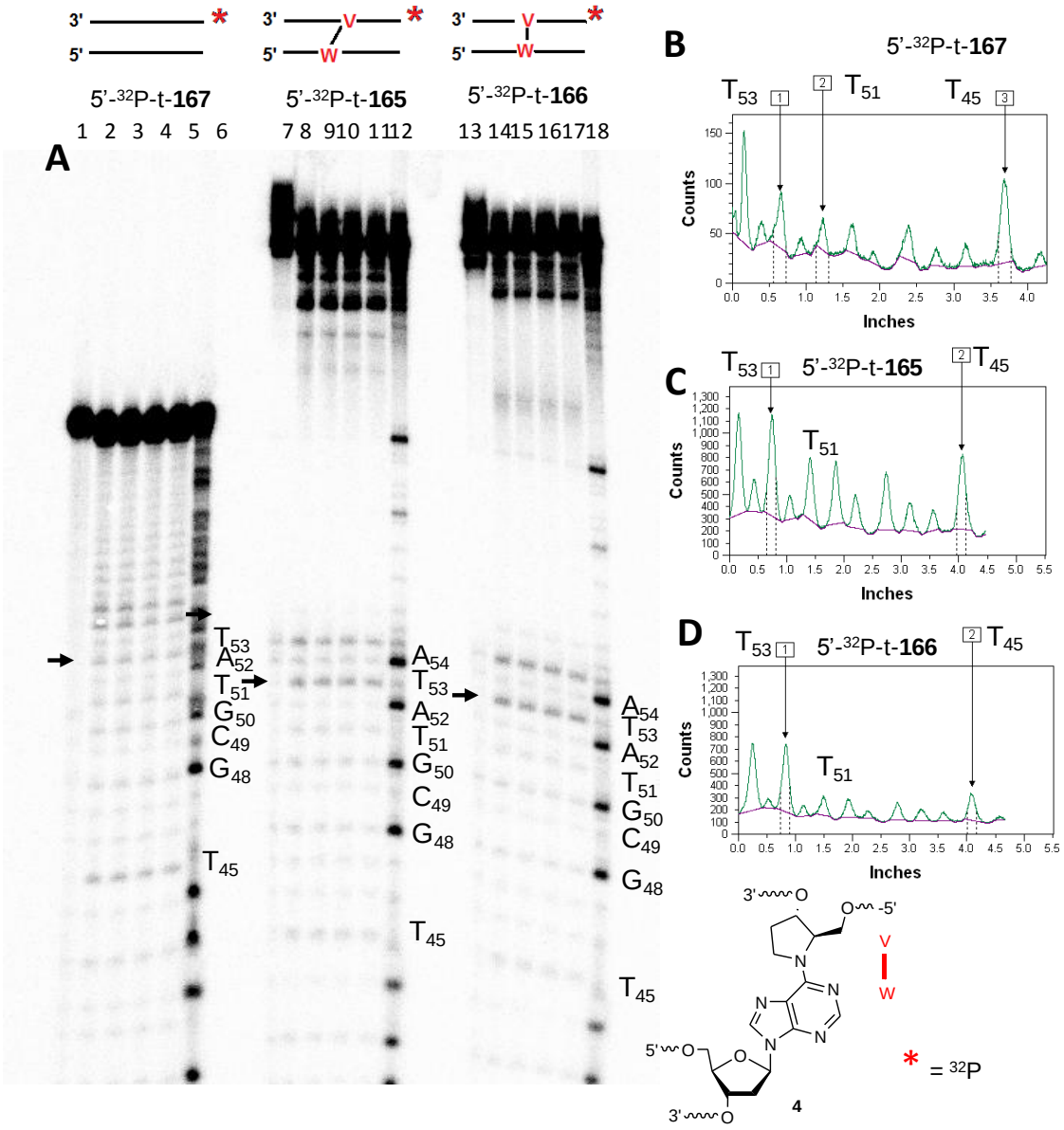
.



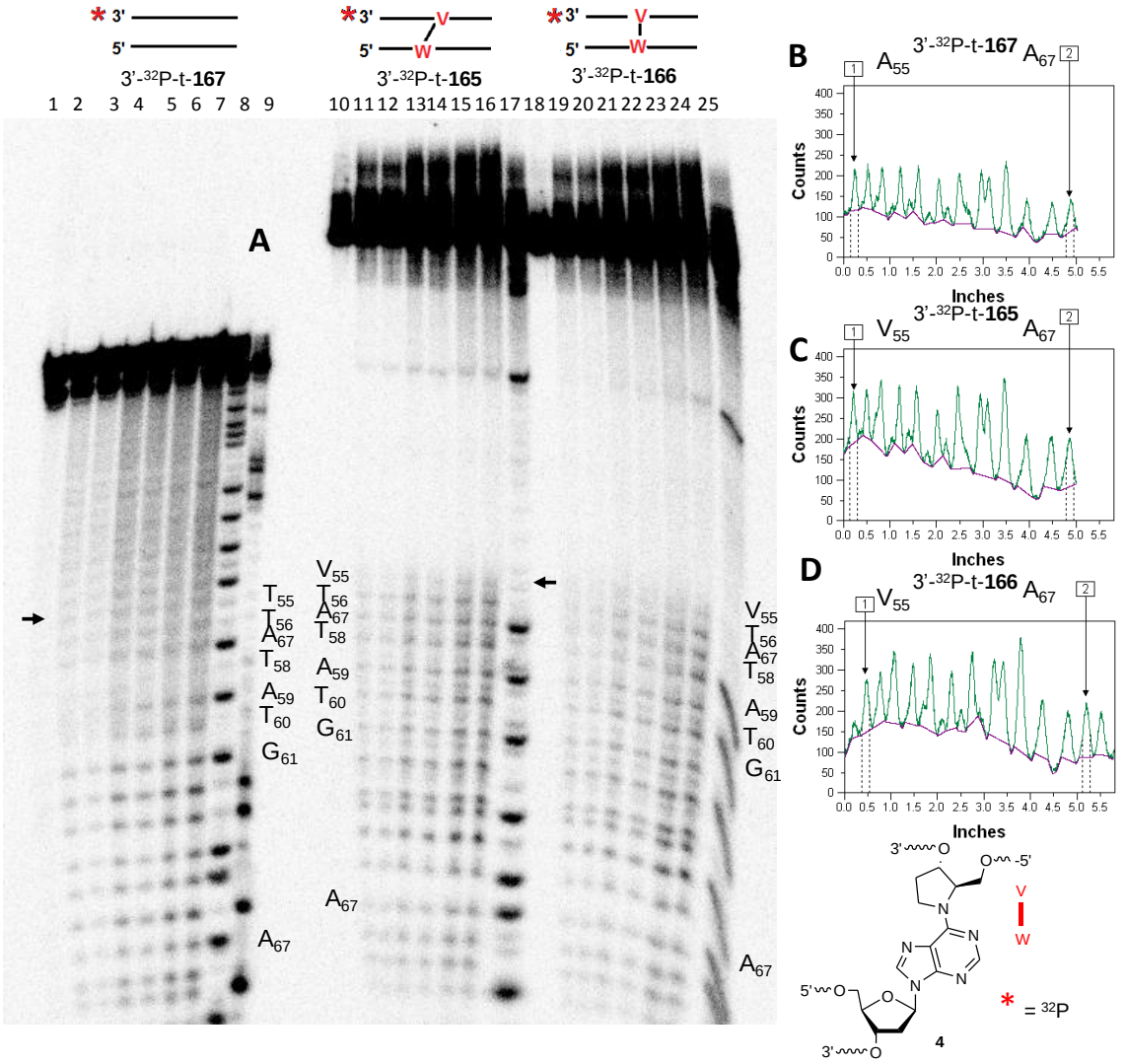
App. Figure 63. DEPC reaction on 5'-^{32}P-t-**167** (control duplex), 5'-^{32}P-t-**165** (observed ICL), and 5'-^{32}P-t-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) DEPC reaction on 5'-^{32}P-t-**167** analyzed on 20% denaturing PAGE. Lanes 1, 1 M piperidine treatment of 5'-^{32}P-t-**167**. Lanes 2 – 9, 3'-^{32}P-t-**167** treated with 1.75 M DEPC for 12 min, room temperature. Lane 10, A-G sequencing of 3'-^{32}P-t-**167**. (B) DEPC reaction on 5'-^{32}P-t-**165** and 5'-^{32}P-t-**166** analyzed on 20% denaturing PAGE. Lanes 1 and 11, 1 M piperidine treatment of 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166**. Lanes 2 – 9 and 12 – 19, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166** treated with 1.75 M DEPC for 12 min, room temperature. Lanes 10 and 20, A-G sequencing of 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166**, respectively.



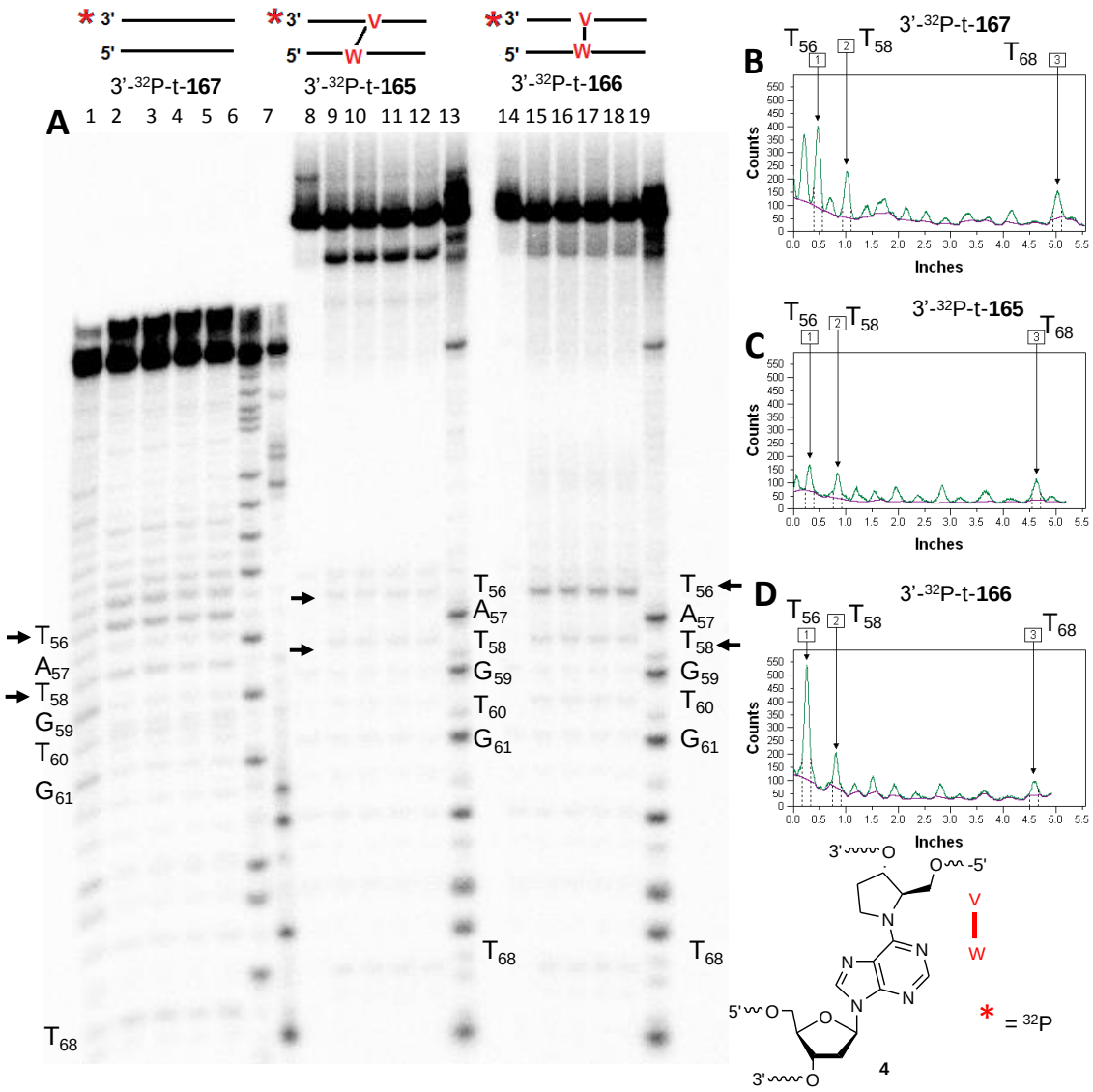
App. Figure 64. Representative Histograms of DEPC reaction on 5'-^{32}P-t-**167** (control duplex), 5'-^{32}P-t-**165** (observed ICL), and 5'-^{32}P-t-**166** (unobserved ICL). Nucleotides of interest are indicated. (A) Representative lane profile from DEPC reaction of 5'-^{32}P-t-**167**. (B) Representative lane profile from DEPC reaction of 5'-^{32}P-t-**165**. (C) Representative lane profile from DEPC reaction of 5'-^{32}P-t-**166**.



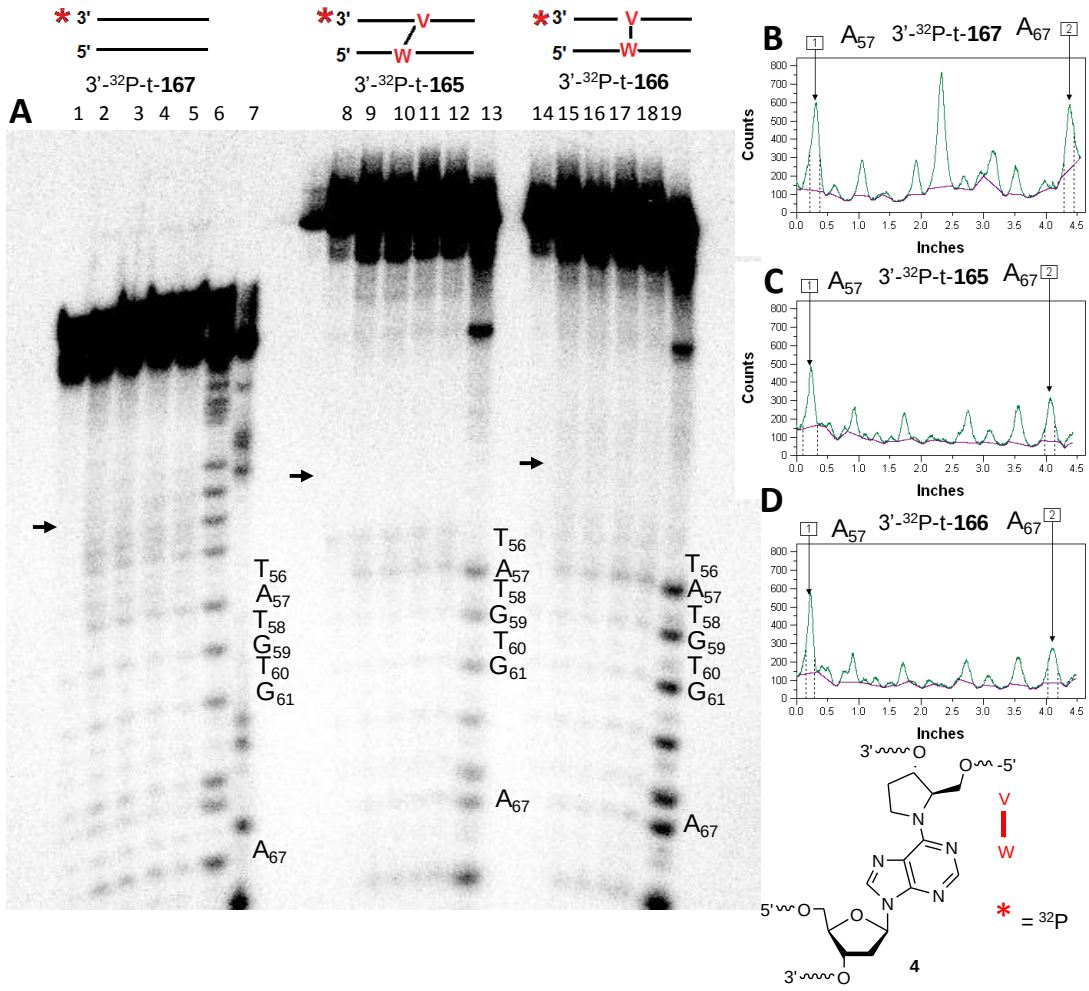
App. Figure 65. $KMnO_4$ reaction on 5'-^{32}P-t-**167** (control duplex), 5'-^{32}P-t-**165** (observed ICL), and 5'-^{32}P-t-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) $KMnO_4$ reaction on 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166** analyzed on 20% denaturing PAGE. Lanes 1, 7, and 13, 1 M piperidine treatment of 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166**. Lanes 2 – 5, 8 – 11 and 14 – 17, 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166** treated with 0.6 mM $KMnO_4$ for 10 min, room temperature. Lanes 6, 12, and 18, A-G sequencing of 5'-^{32}P-t-**167**, 5'-^{32}P-t-**165**, and 5'-^{32}P-t-**166**, respectively. (B) Representative lane profile from $KMnO_4$ reaction of 5'-^{32}P-t-**167**. (C) Representative lane profile from $KMnO_4$ reaction of 5'-^{32}P-t-**165**. (D) Representative lane profile from $KMnO_4$ reaction of 5'-^{32}P-t-**166**.



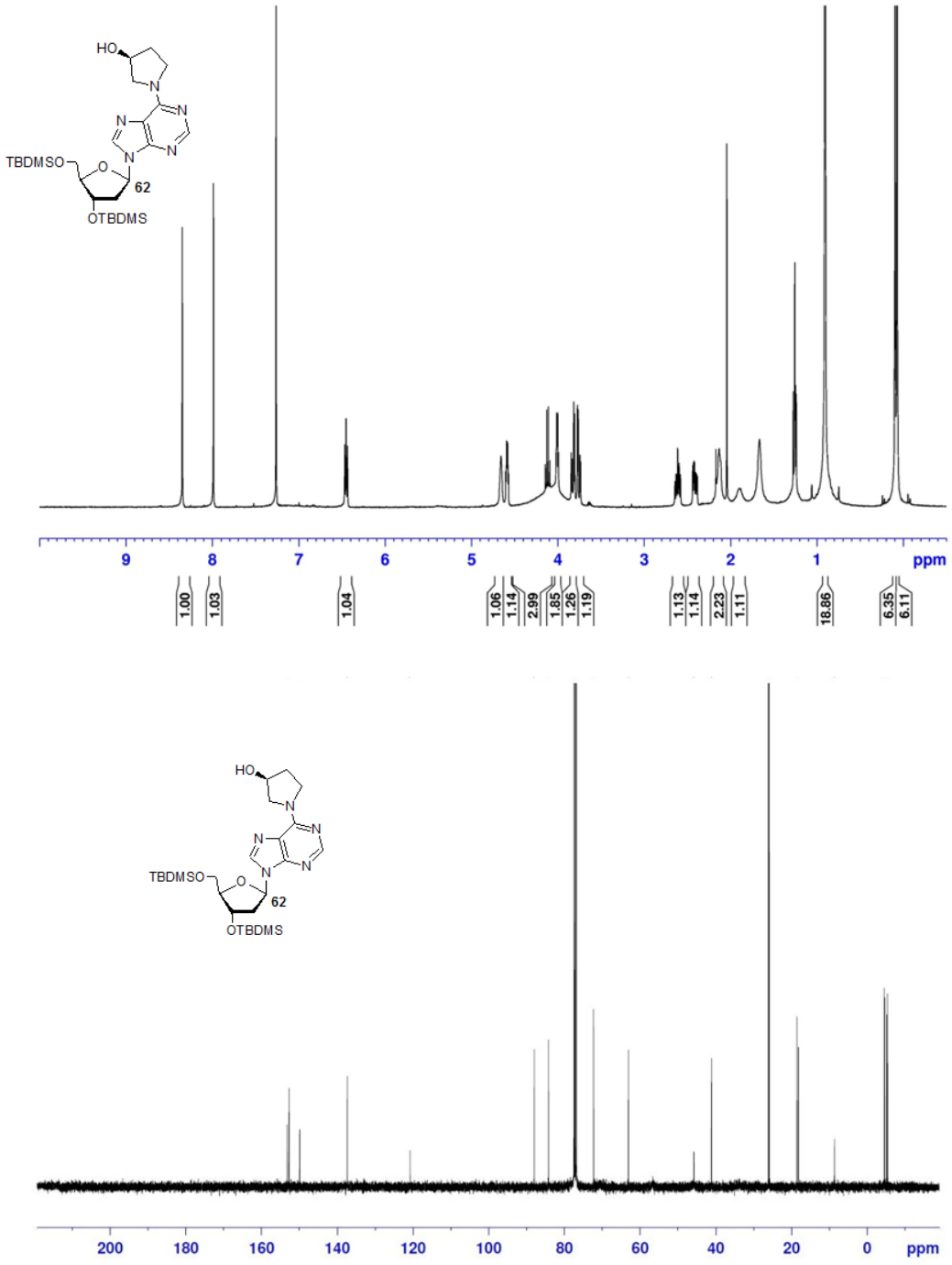
App. Figure 66. Hydroxyl radical cleavage reaction on 3'-^{32}P-t-**167** (control duplex), 3'-^{32}P-t-**165** (observed ICL), and 3'-^{32}P-t-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) Hydroxyl radical cleavage reaction on 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** analyzed on 20% denaturing PAGE. Lanes 1, 10 and 18, 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** without any treatment. Lanes 2 – 7, 11 – 16 and 19 – 24, 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** subjected to hydroxyl radical cleavage reaction for 5 min, room temperature. Lanes 8 and 9, A-G and T-C sequencing of 3'-^{32}P-t-**167**, respectively. Lanes 17 and 25, A-G sequencing of 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166**, respectively. (B) Representative lane profile from hydroxyl radical cleavage reaction of 3'-^{32}P-t-**167**. (C) Representative lane profile from hydroxyl radical cleavage reaction of 3'-^{32}P-t-**165**. (D) Representative lane profile from hydroxyl radical cleavage reaction of 3'-^{32}P-t-**166**.



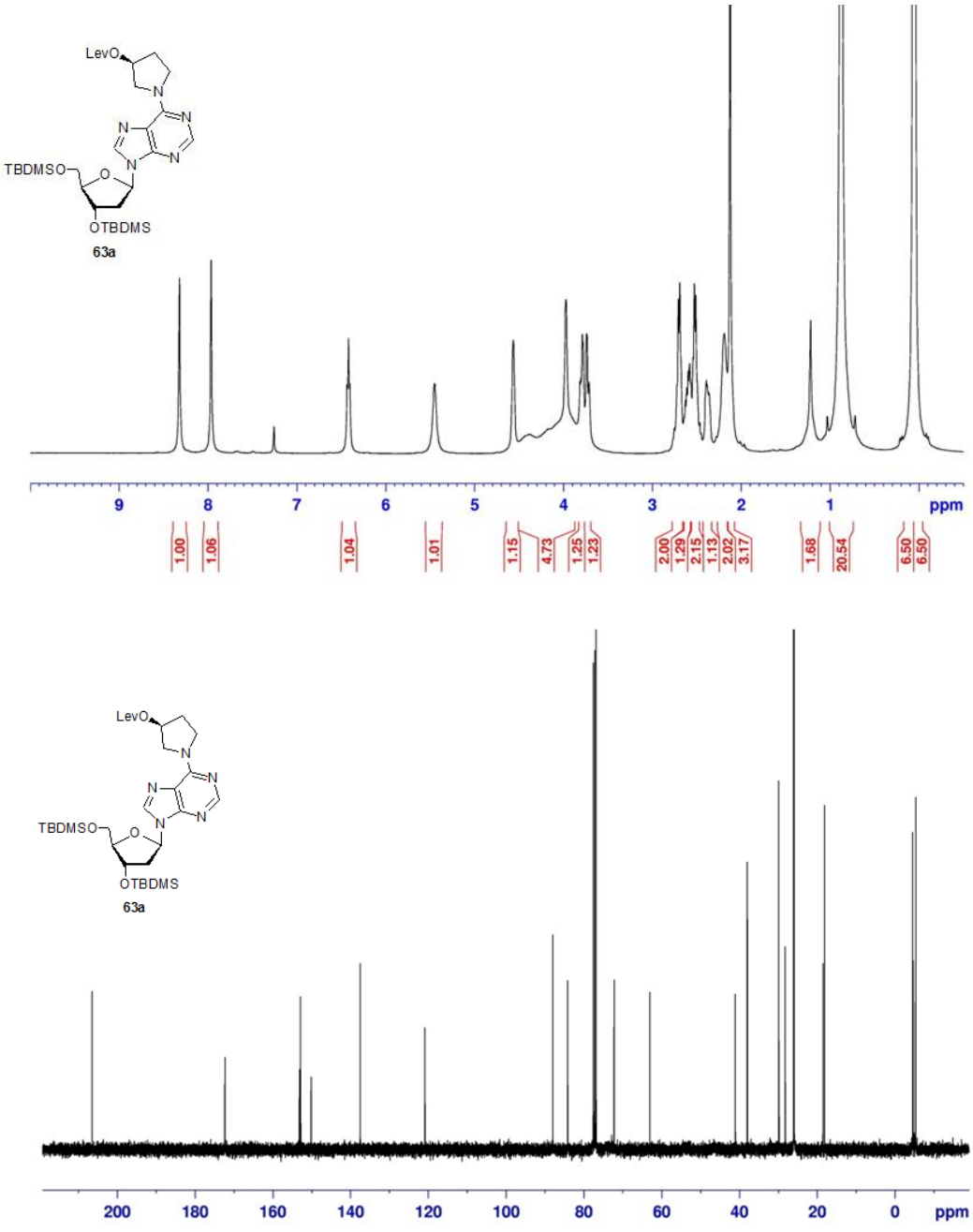
App. Figure 67. $KMnO_4$ reaction on 3'-^{32}P-t-**167** (control duplex), 3'-^{32}P-t-**165** (observed ICL), and 3'-^{32}P-t-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) $KMnO_4$ reaction on 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** analyzed on 20% denaturing PAGE. Lanes 1, 8, and 14, 1 M piperidine treatment of 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166**. Lanes 2 – 5, 9 – 12 and 15 – 18, 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** treated with 0.6 mM $KMnO_4$ for 10 min, room temperature. Lanes 6 and 7, A-G and T-C sequencing of 3'-^{32}P-t-**167**, respectively. Lanes 13 and 19, A-G sequencing of 3'-^{32}P-t-**165** and 3'-^{32}P-t-**166**, respectively. (B) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**167**. (C) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**165**. (D) Representative lane profile from $KMnO_4$ reaction of 3'-^{32}P-t-**166**.



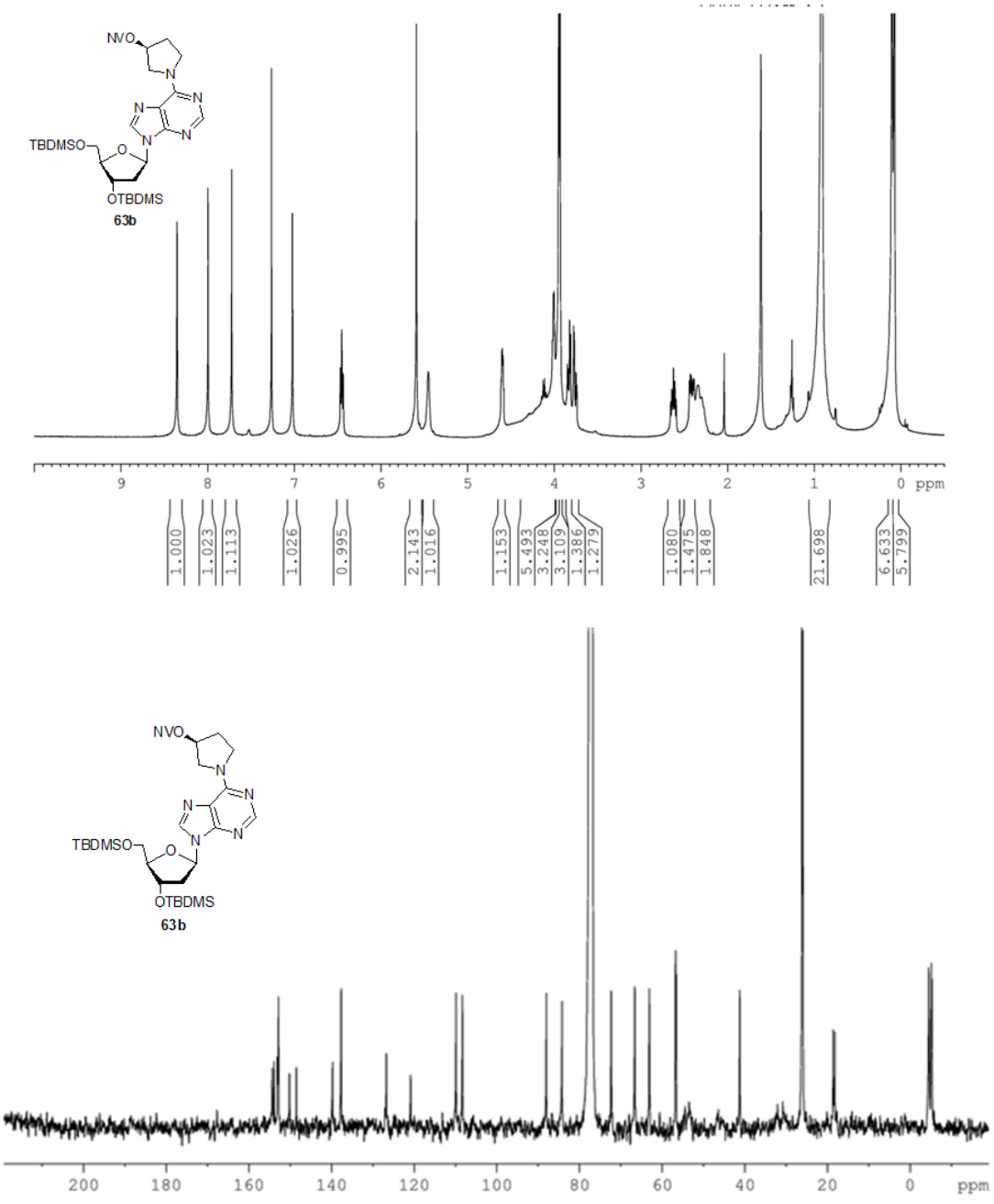
App. Figure 68. DEPC reaction on 3'-^{32}P-t-**167** (control duplex), 3'-^{32}P-t-**165** (observed ICL), and 3'-^{32}P-t-**166** (unobserved ICL). Nucleotides of interest are indicated with arrows. (A) DEPC reaction on 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** analyzed on 20% denaturing PAGE. Lanes 1, 8, and 14, 1 M piperidine treatment of 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166**. Lanes 2 – 5, 9 – 12 and 15 – 18, 3'-^{32}P-t-**167**, 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166** treated with 1.75 M DEPC for 12 min, room temperature. Lanes 6 and 7, A-G and T-C sequencing of 3'-^{32}P-t-**167**, respectively. Lanes 13 and 19, A-G sequencing of 3'-^{32}P-t-**165**, and 3'-^{32}P-t-**166**, respectively. (C) Representative lane profile from DEPC reaction of 3'-^{32}P-t-**167**. (D) Representative lane profile from DEPC reaction of 3'-^{32}P-t-**165**. (E) Representative lane profile from DEPC reaction of 3'-^{32}P-t-**166**.



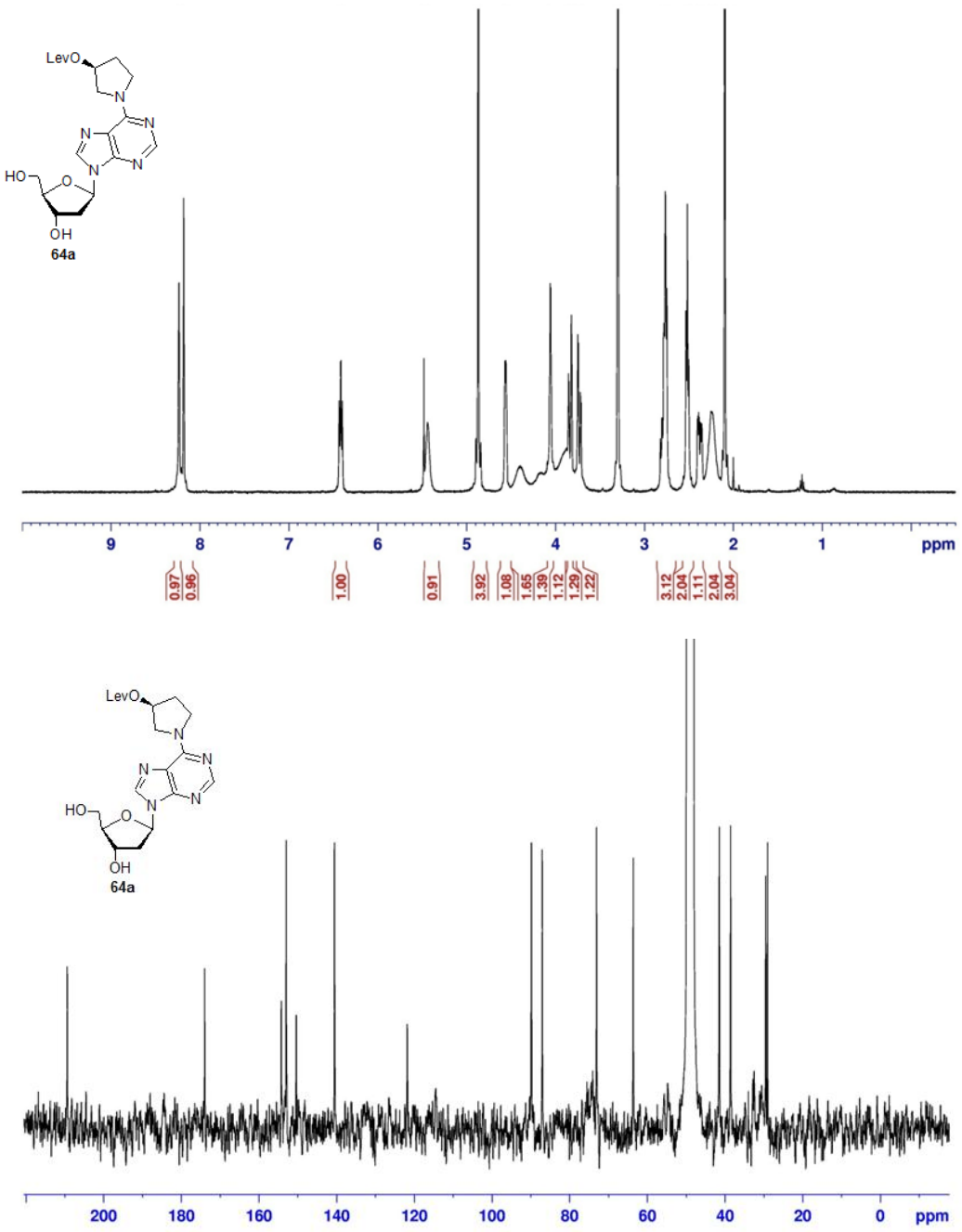
App. Figure 69. ^{1}H and ^{13}C NMR spectra of **62.**



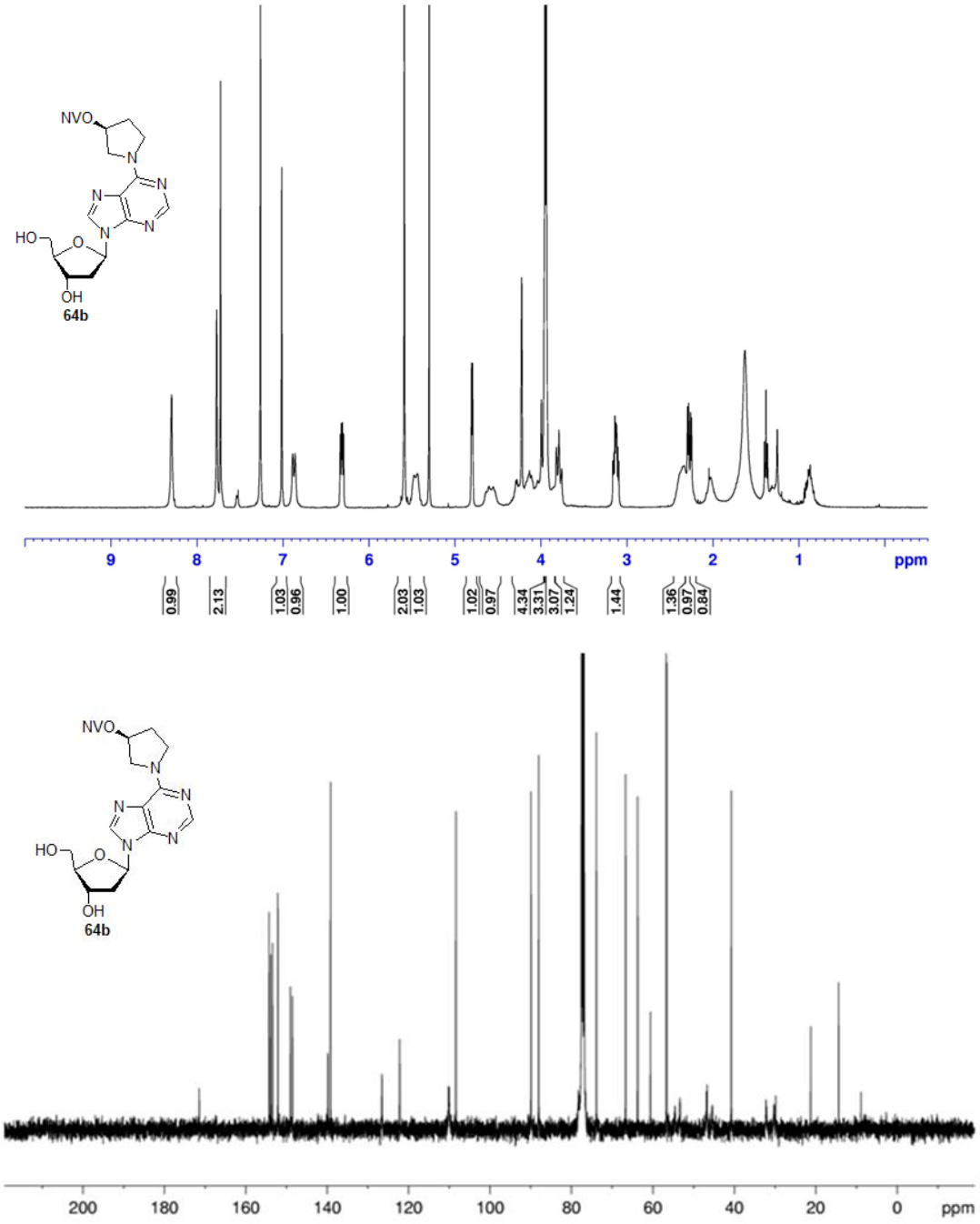
App. Figure 70. ^{1}H and ^{13}C NMR spectra of **63a.**



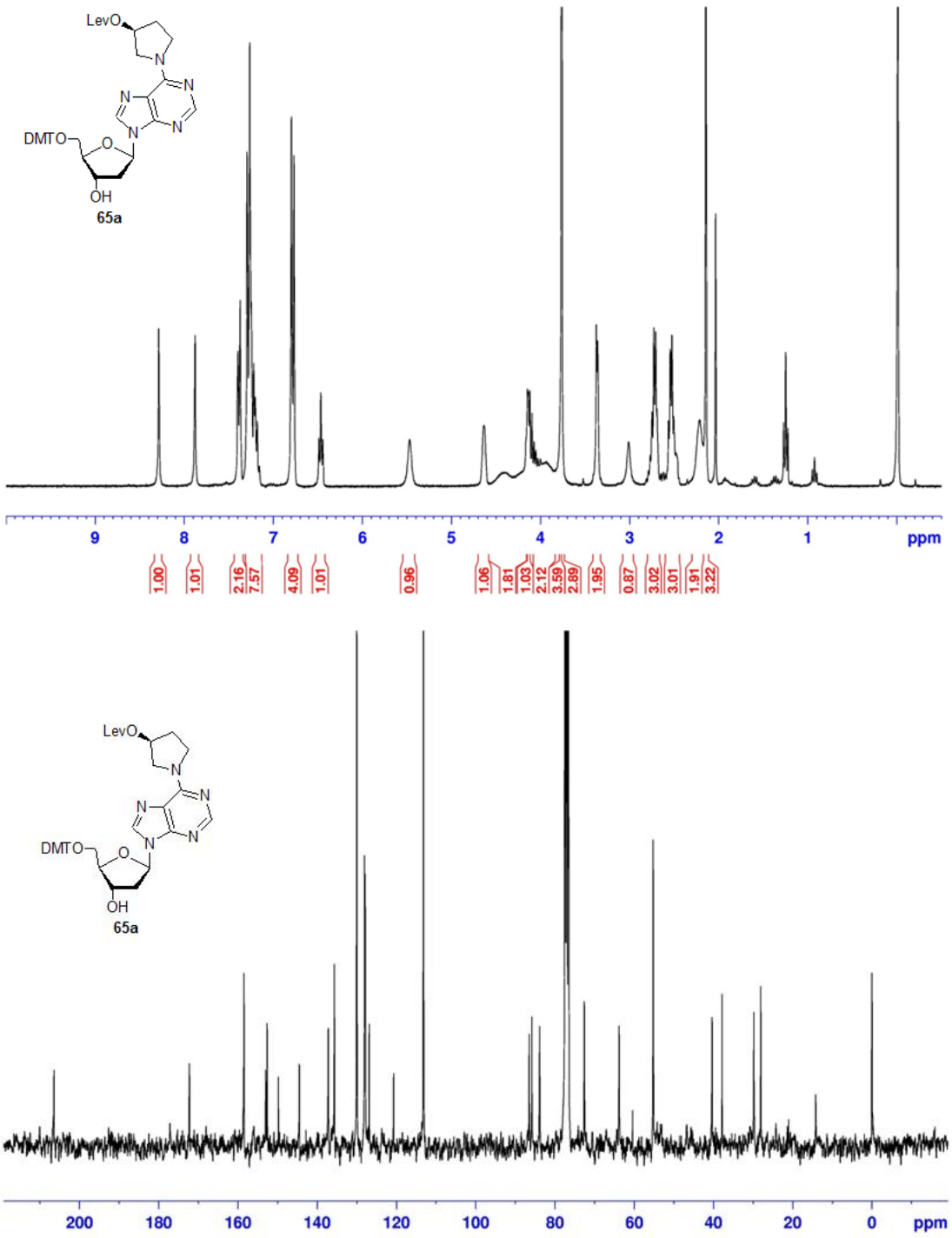
App. Figure 71. ^{1}H and ^{13}C NMR spectra of **63b.**



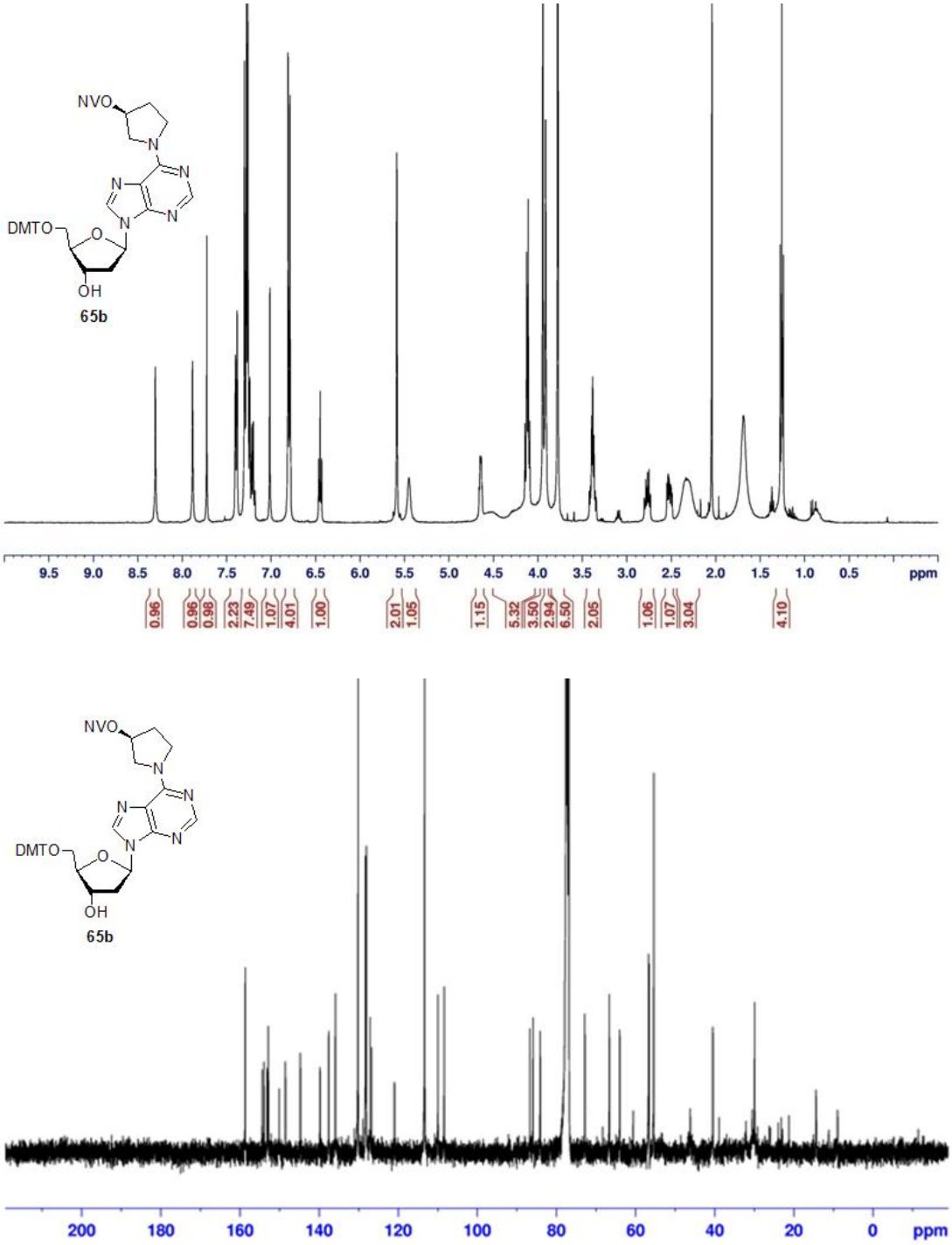
App. Figure 72. ^{1}H and ^{13}C NMR spectra of **64a.**



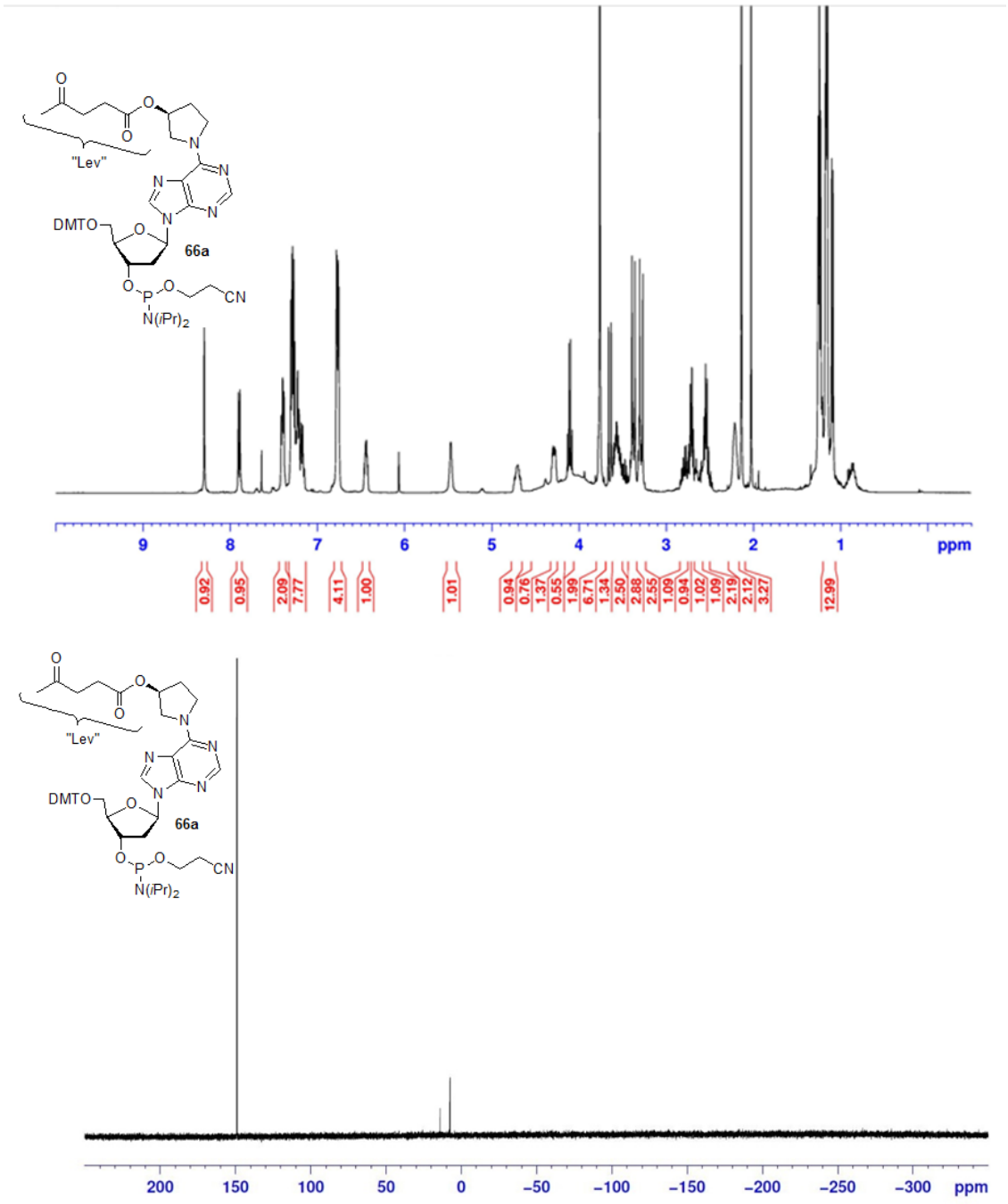
App. Figure 73. ^{1}H and ^{13}C NMR spectra of **64b.**



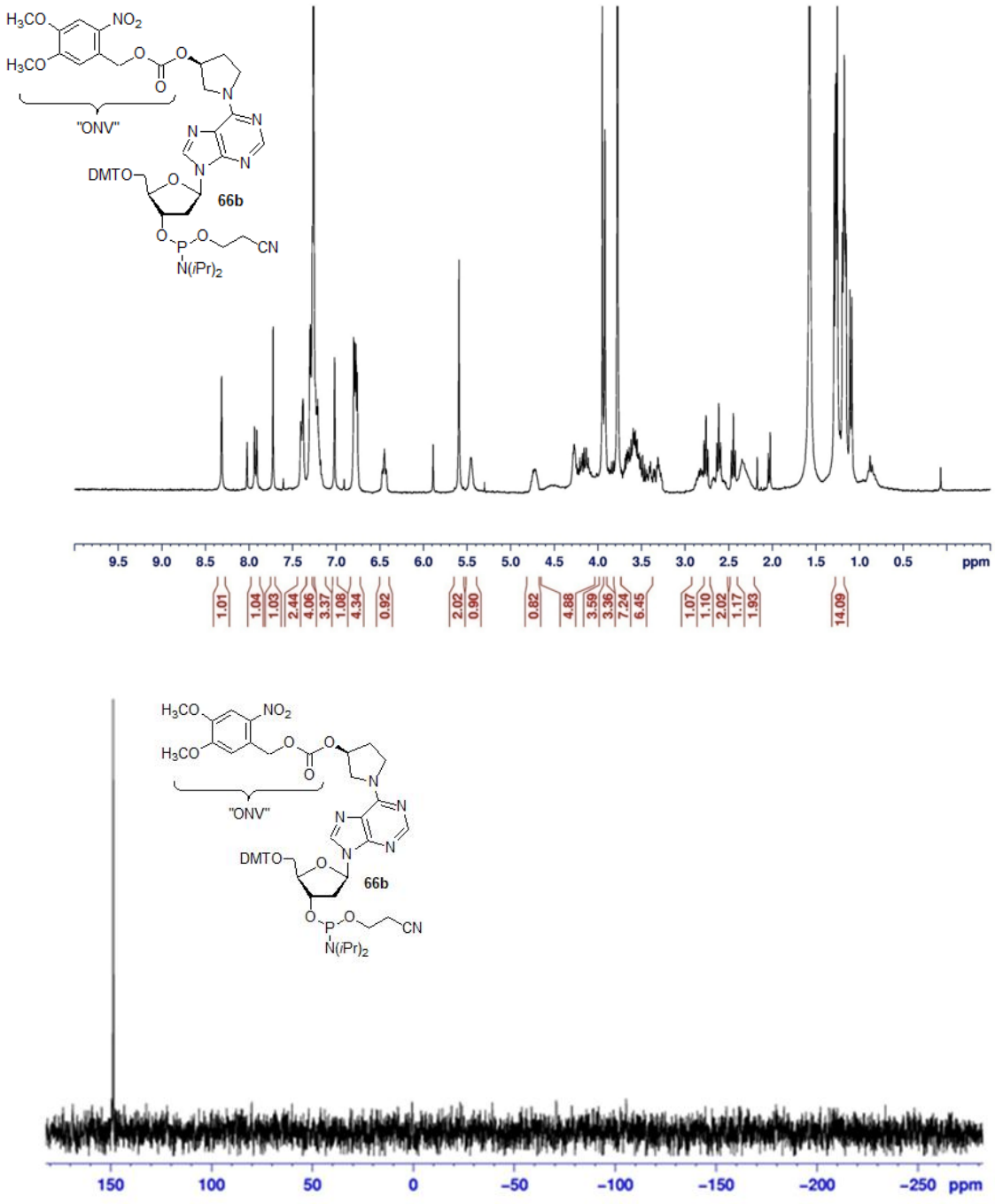
App. Figure 74. ^{1}H and ^{13}C NMR spectra of **65a.**



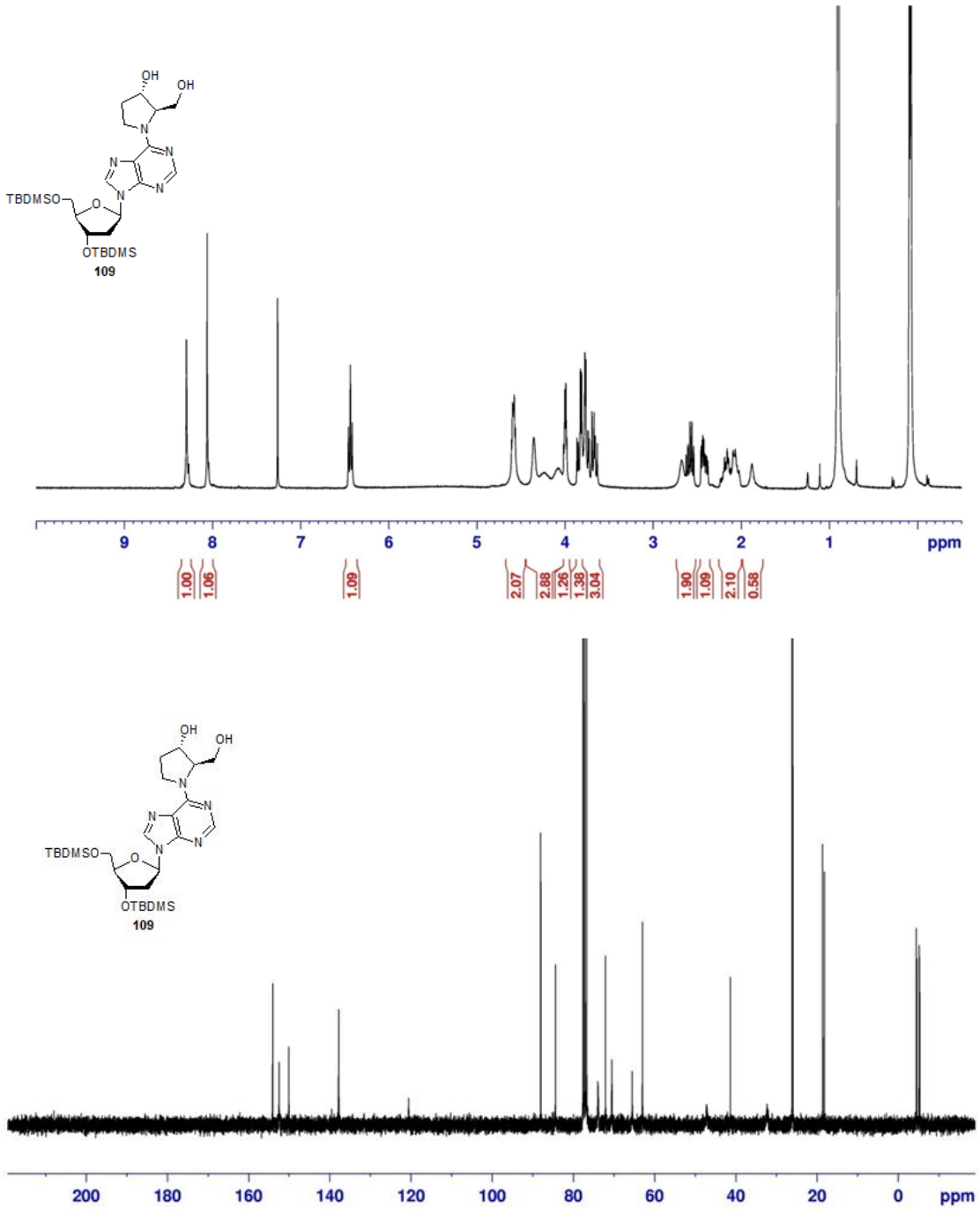
App. Figure 75. ^{1}H and ^{13}C NMR spectra of **65b.**



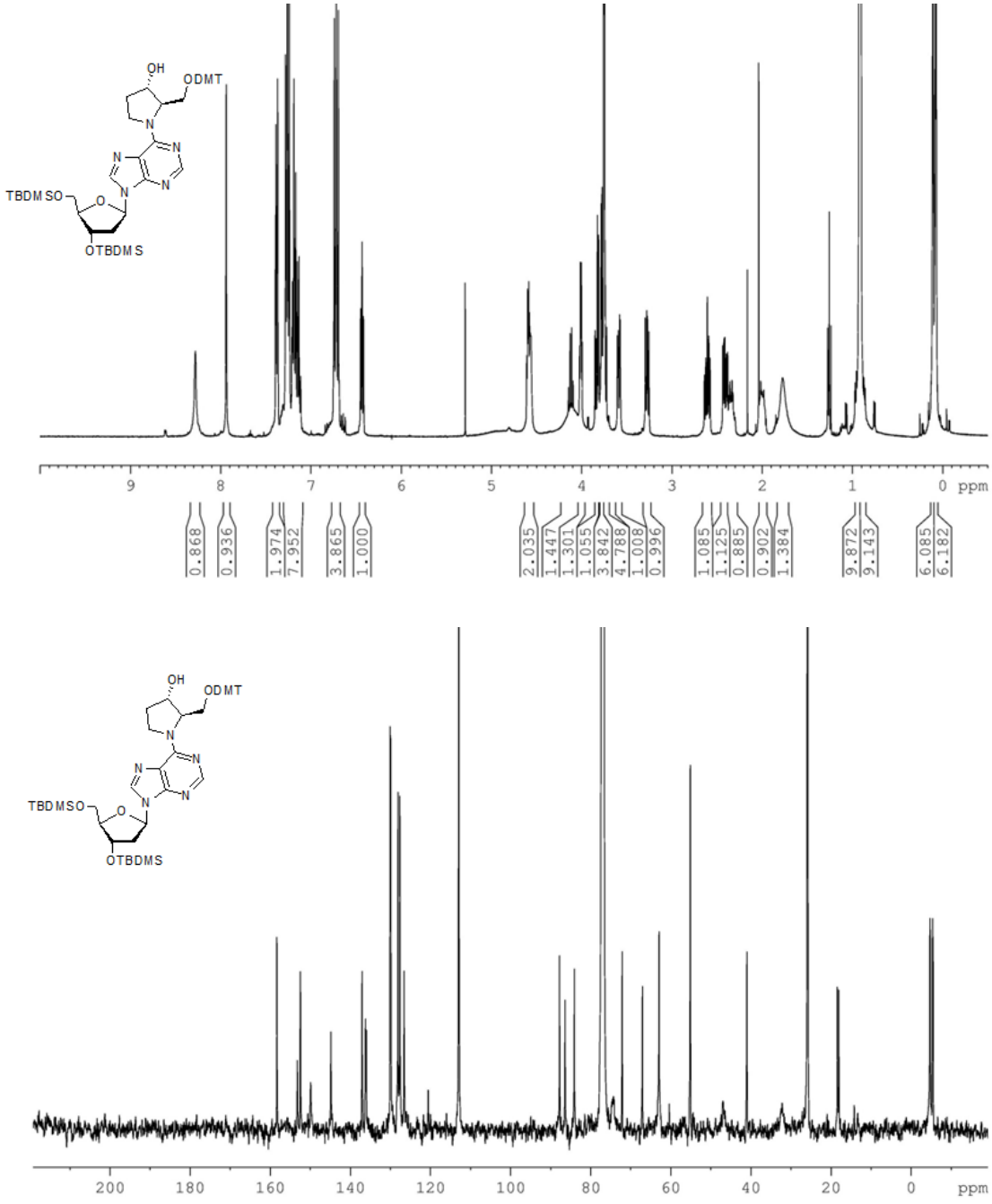
App. Figure 76. ^{1}H and ^{13}C NMR spectra of **66a.**



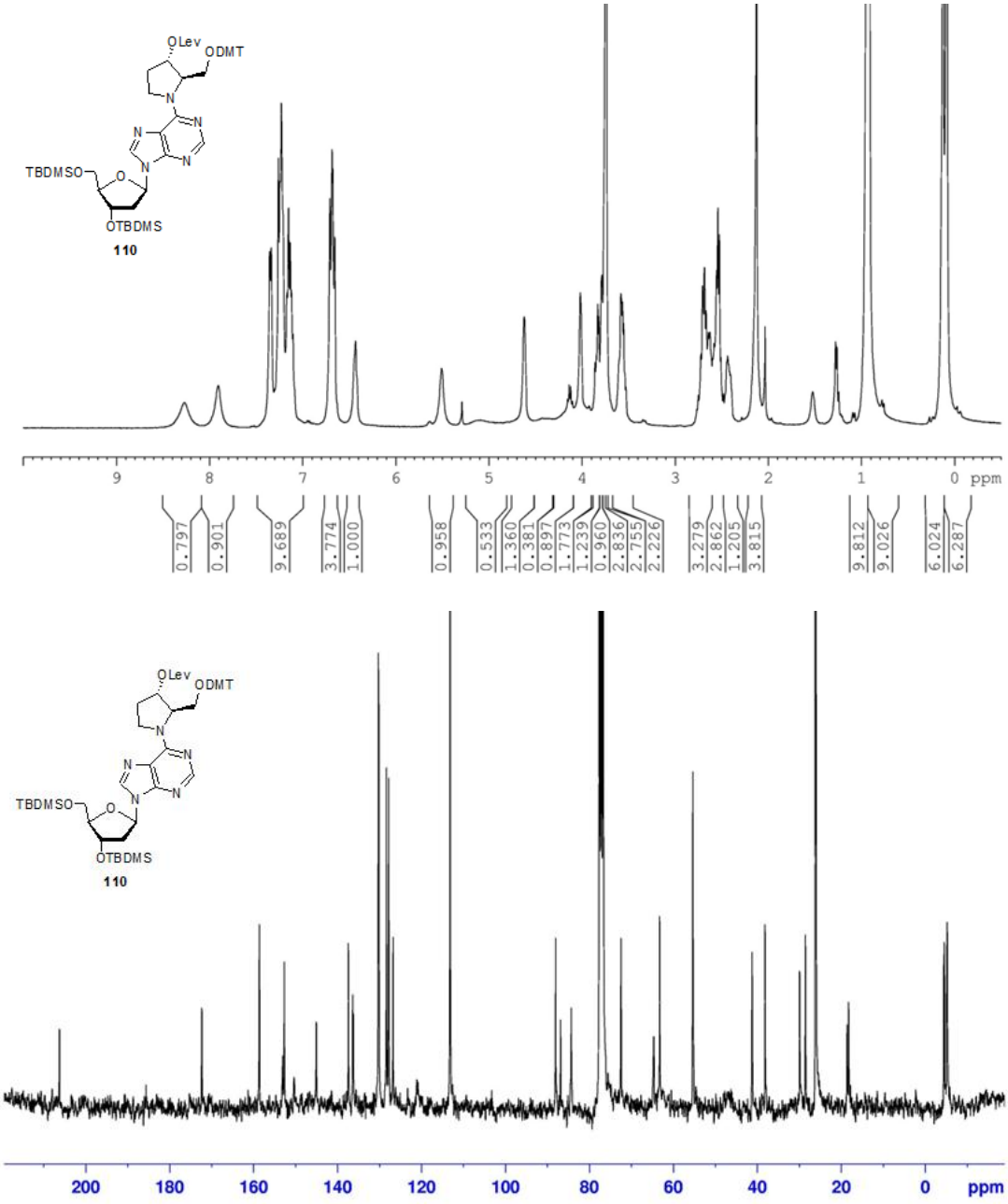
App. Figure 77. ^{1}H and ^{13}C NMR spectra of **66b.**



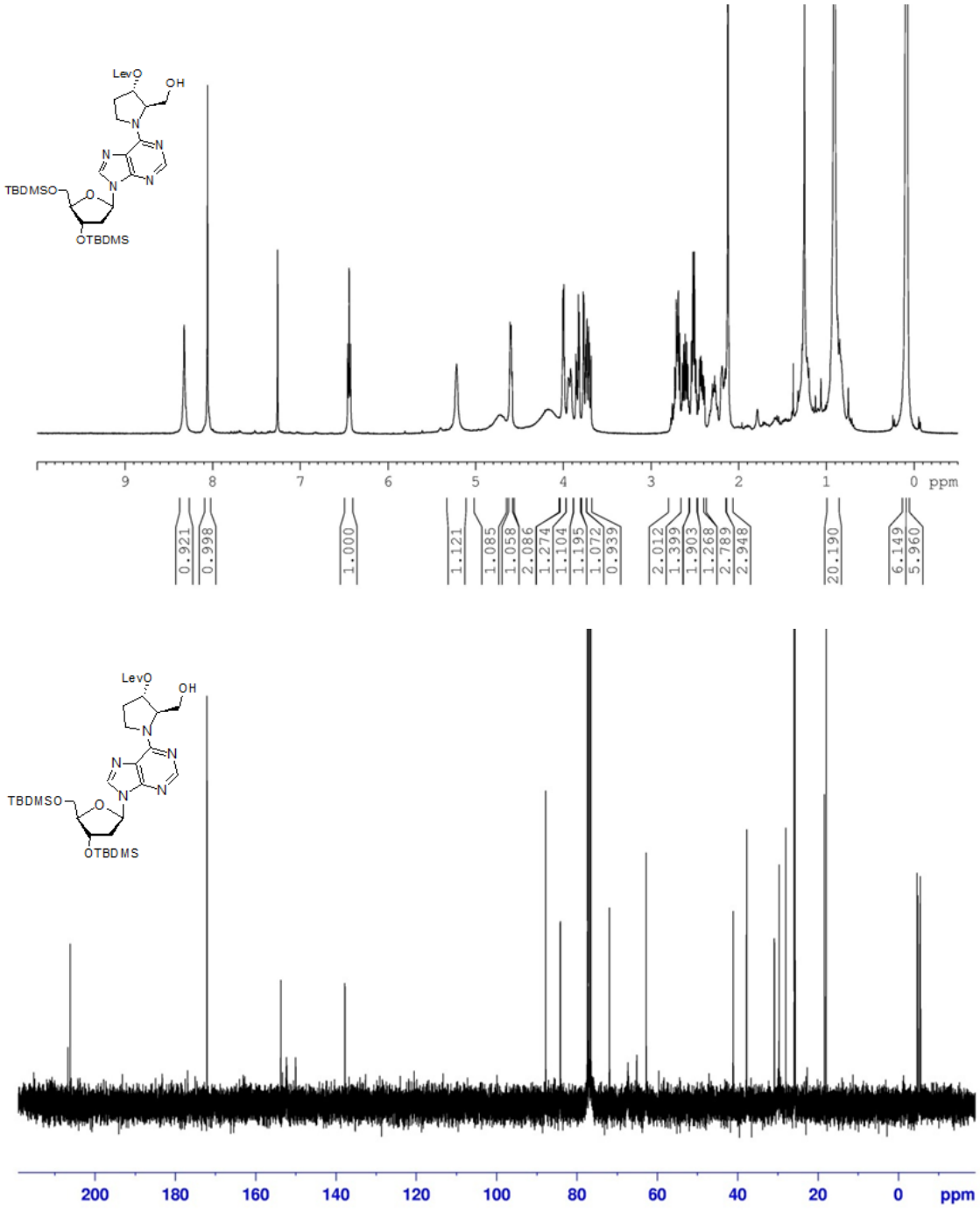
App. Figure 78. ^{1}H and ^{13}C NMR spectra of **109.**



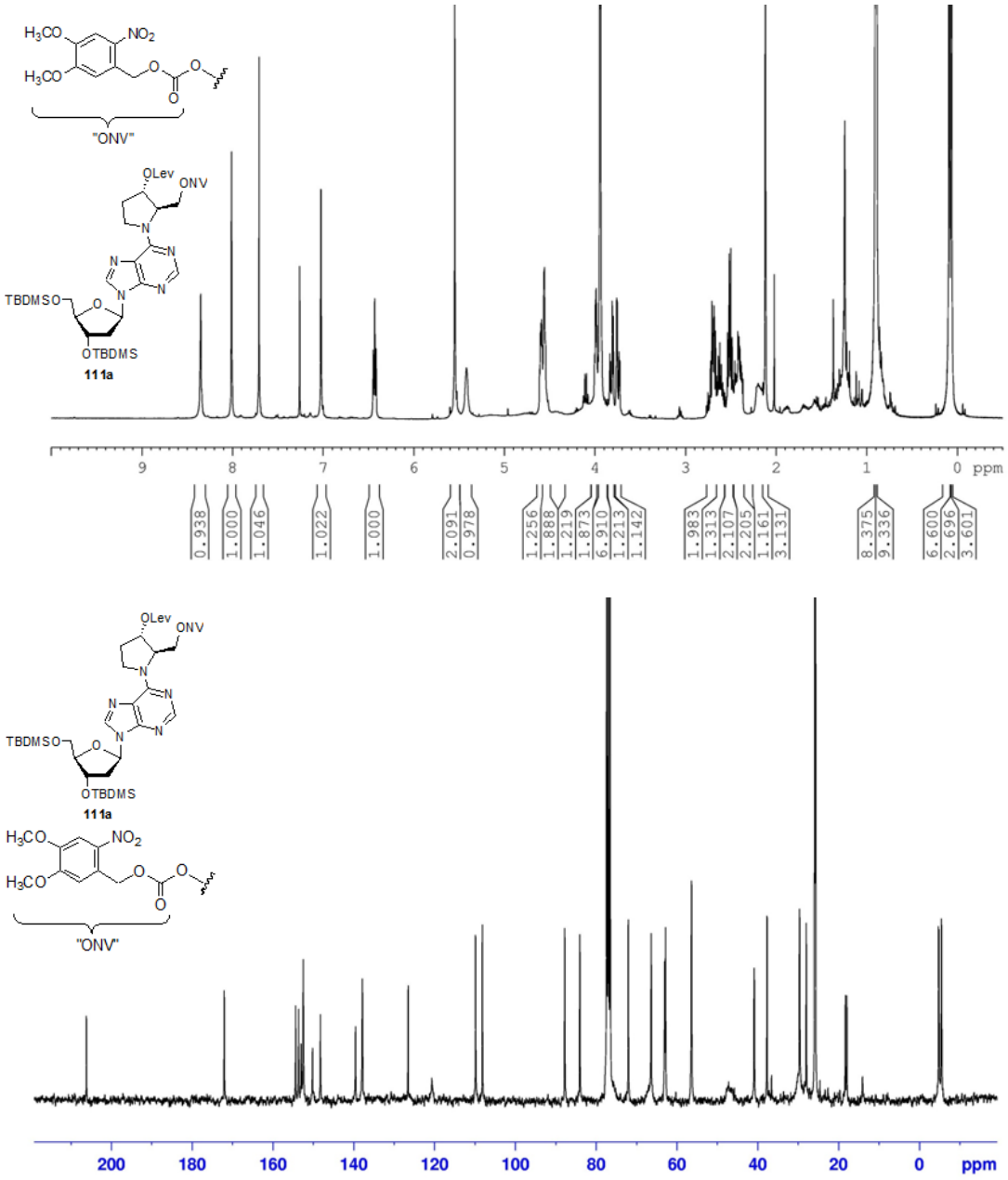
App. Figure 79. ^{1}H and ^{13}C NMR spectra of DMT protected **109.**



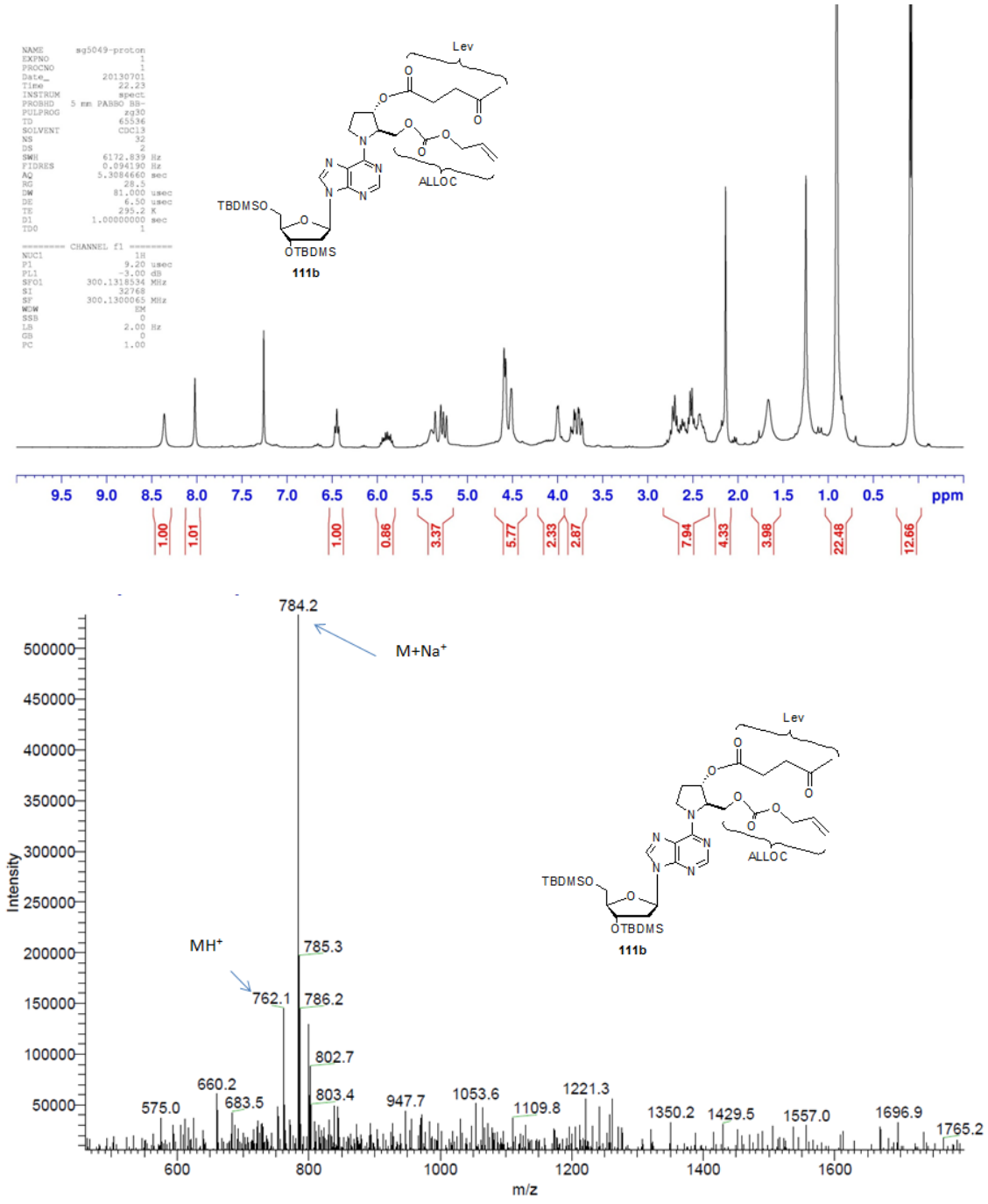
App. Figure 80. ^{1}H and ^{13}C NMR spectra of **110.**



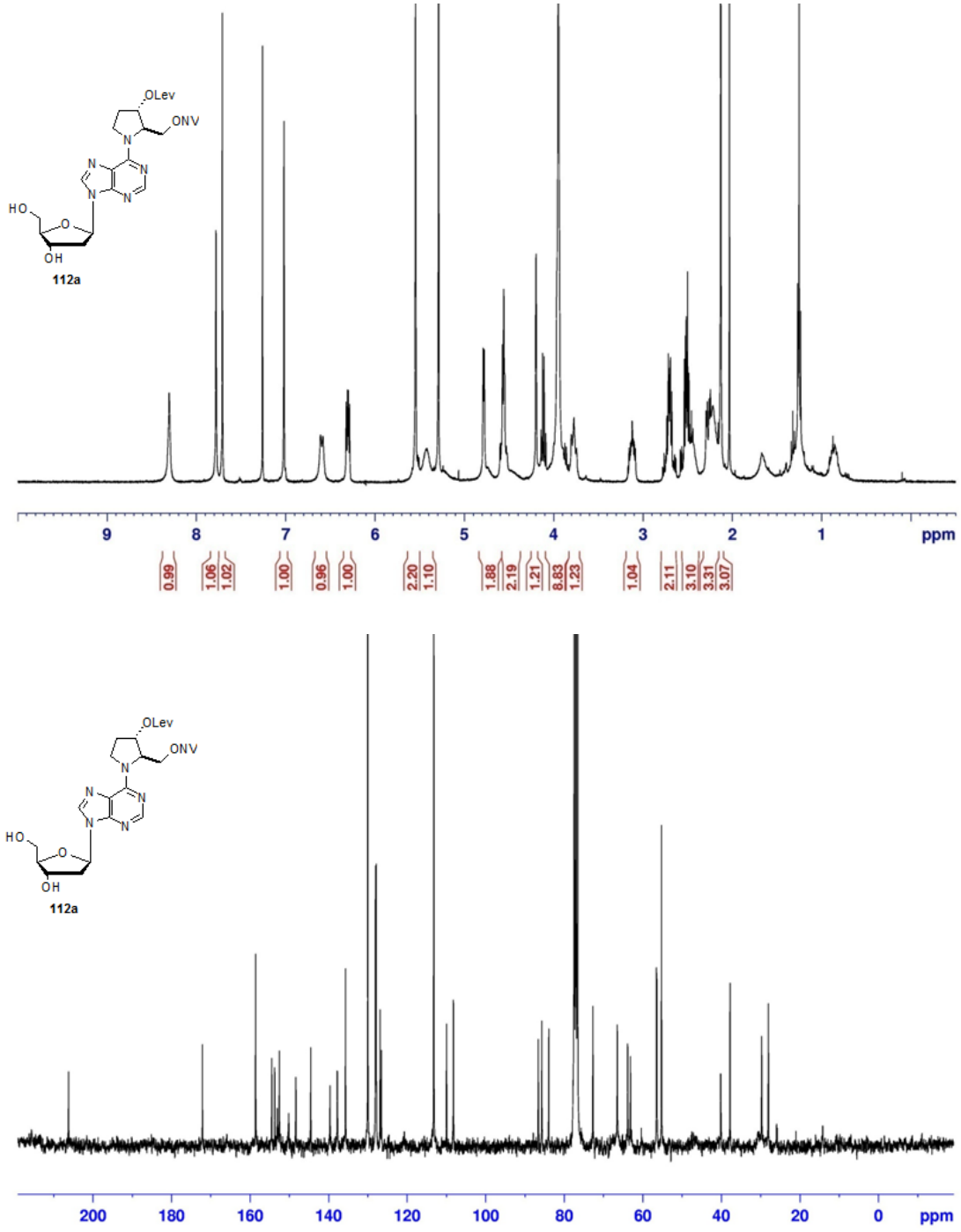
App. Figure 81. ^{1}H and ^{13}C NMR spectra of DMT deprotected **110.**



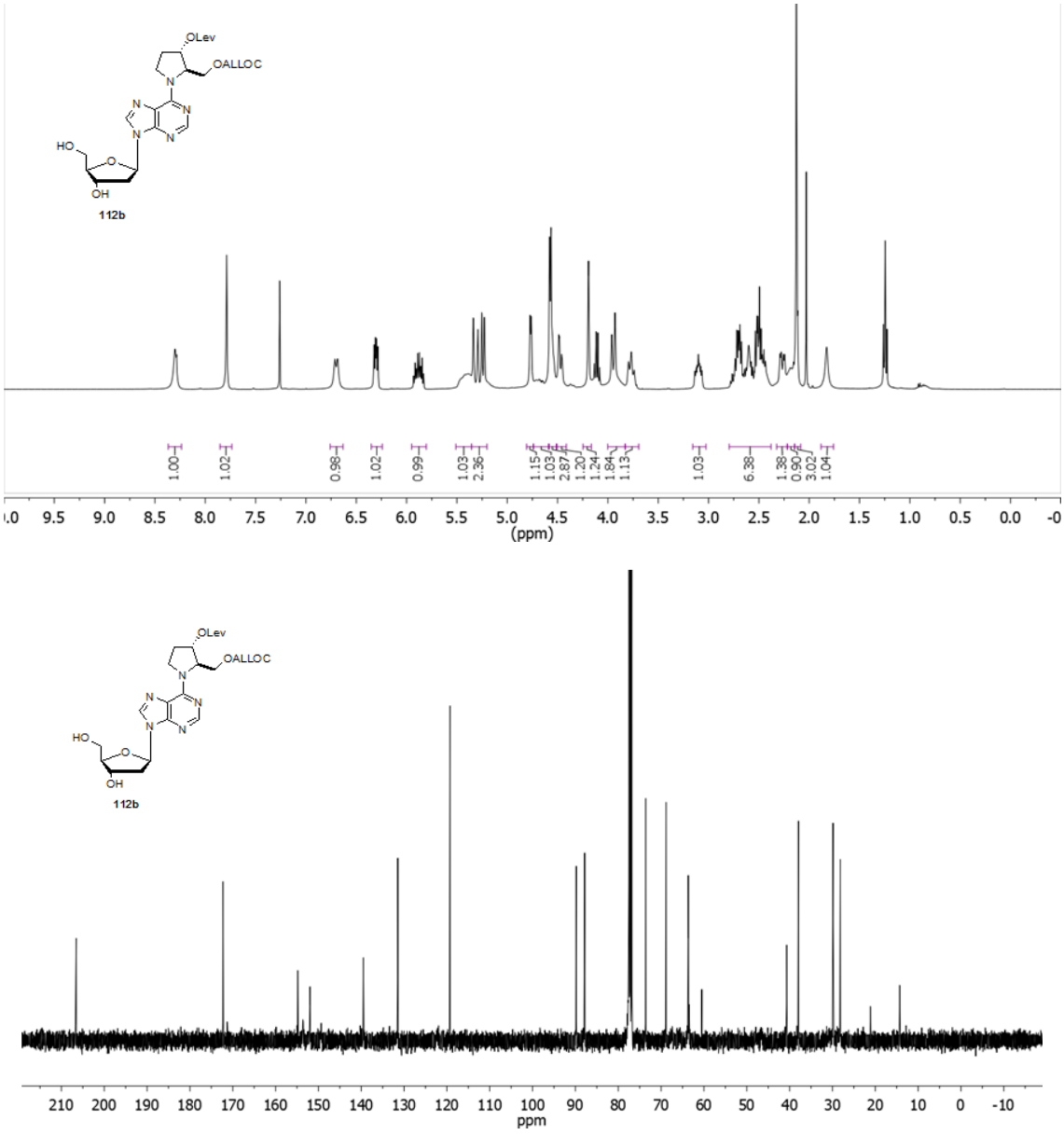
App. Figure 82. ^{1}H and ^{13}C NMR spectra of **111a.**



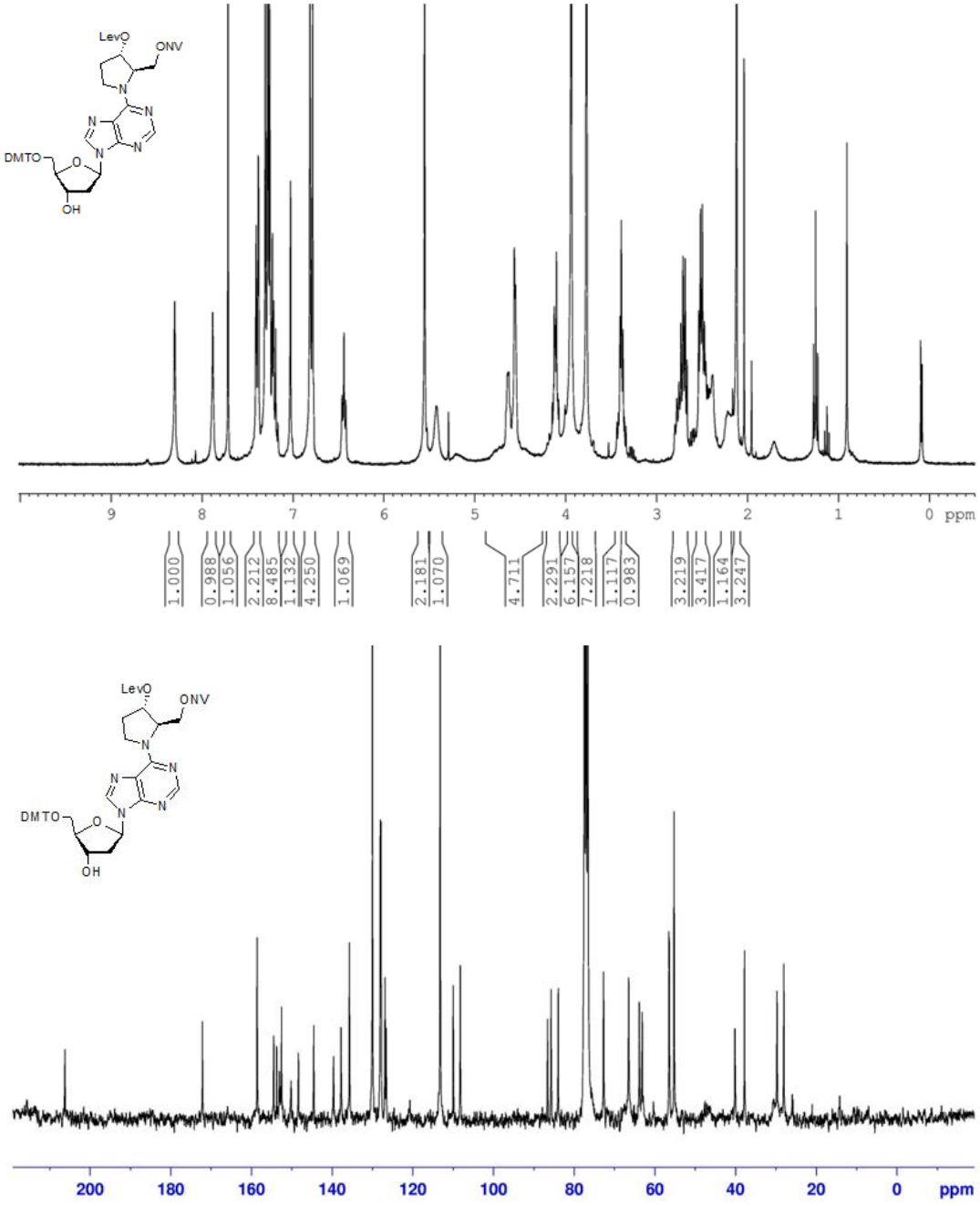
App. Figure 83. ^{1}H NMR and ESI-MS spectra of **111b.**



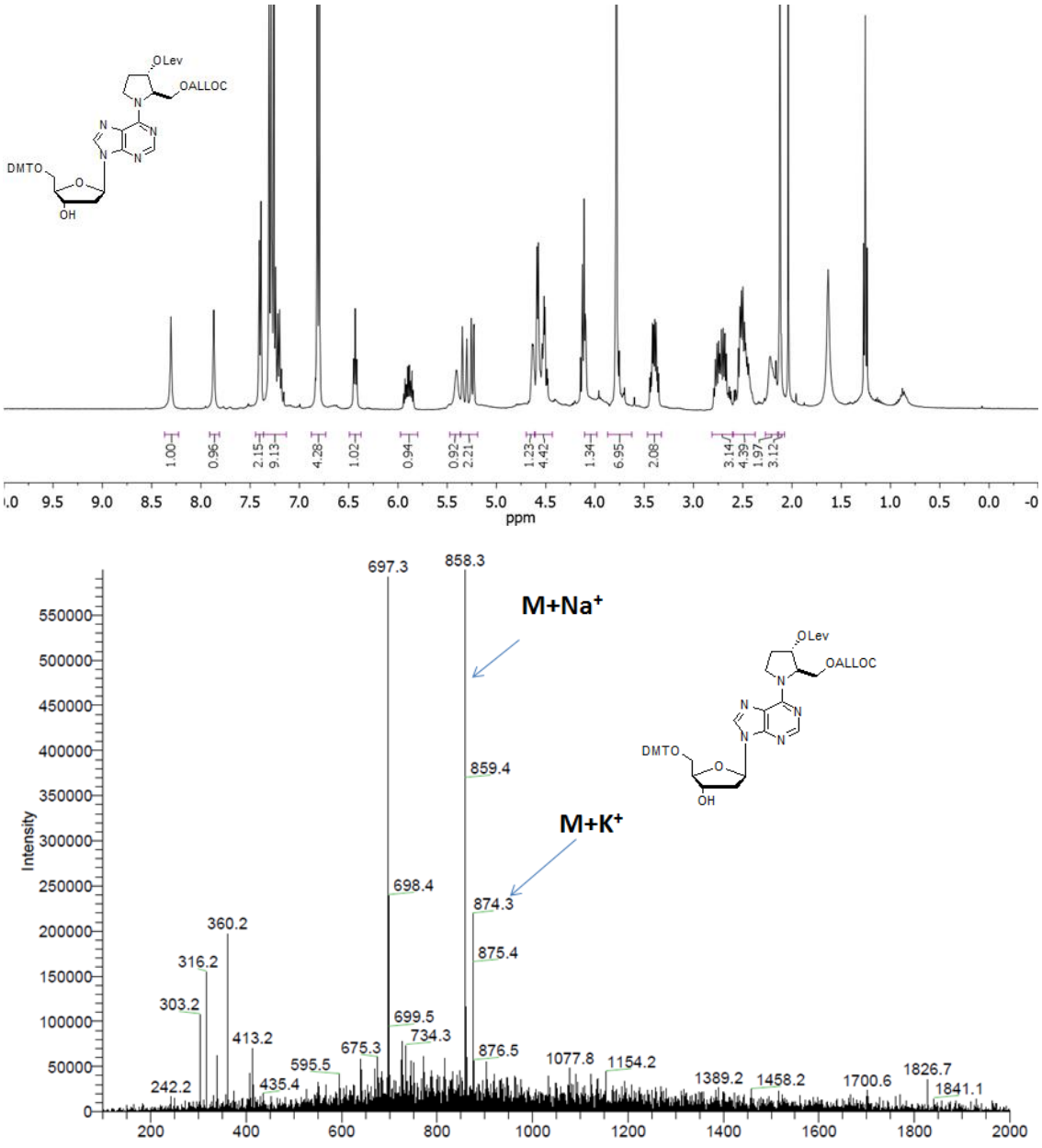
App. Figure 84. ^{1}H and ^{13}C NMR spectra of **112a.**



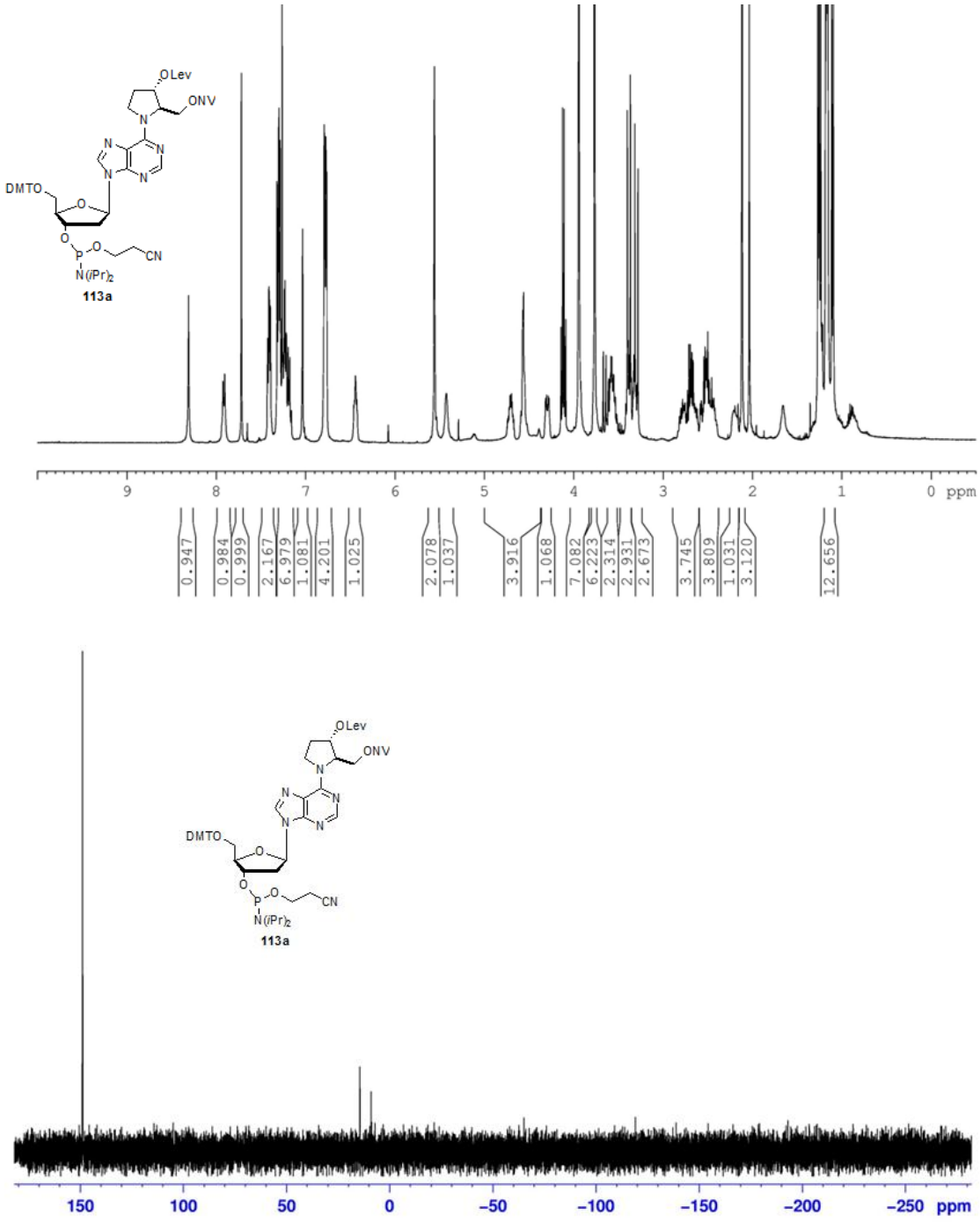
App. Figure 85. ^{1}H and ^{13}C NMR spectra of **112b.**



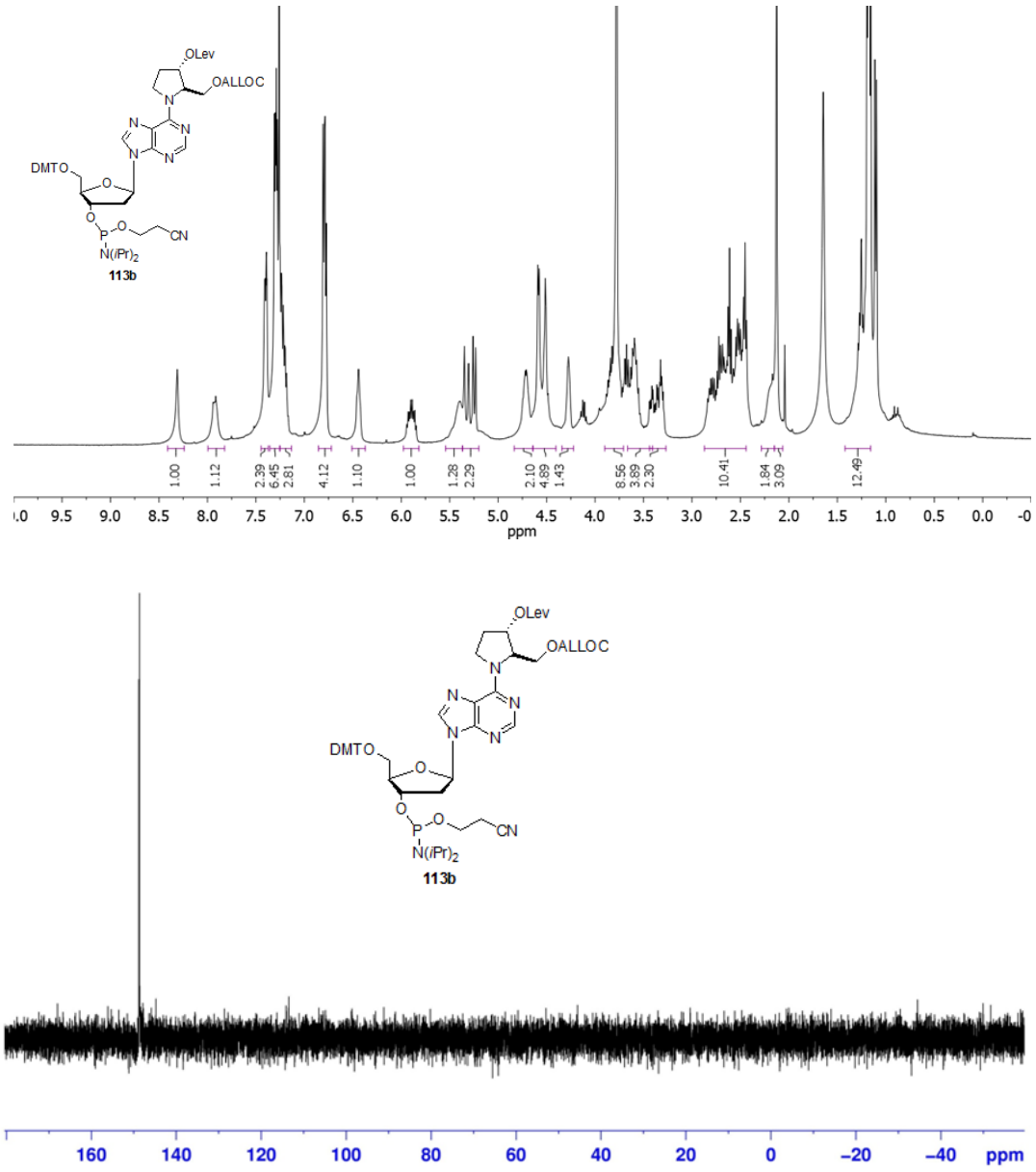
App. Figure 86. ^{1}H and ^{13}C NMR spectra of DMT protected **112a.**



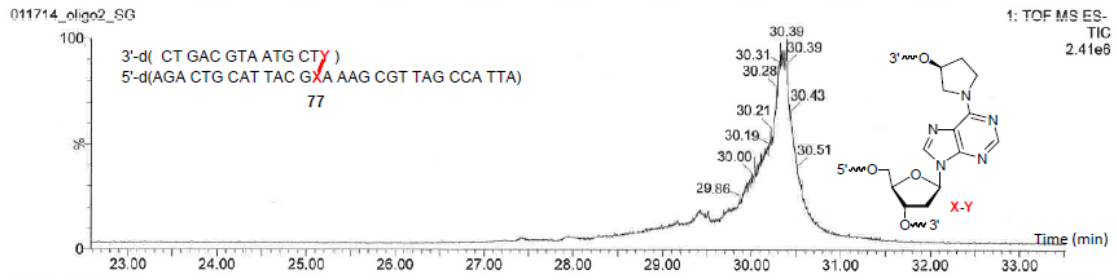
App. Figure 87. ^{1}H NMR and ESI-MS spectra of DMT protected **112b.**



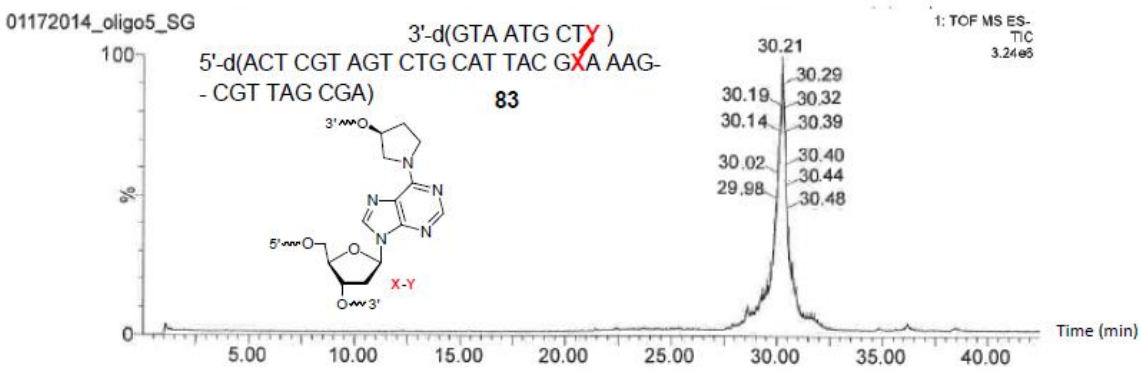
App. Figure 88. ^{1}H and ^{31}P NMR spectra **113a.**



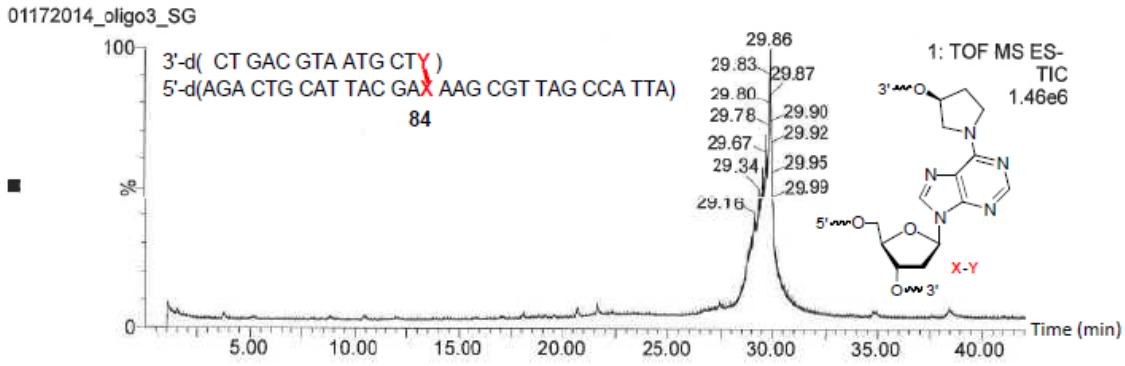
App. Figure 89. ^{1}H and ^{31}P NMR spectra **113b.**



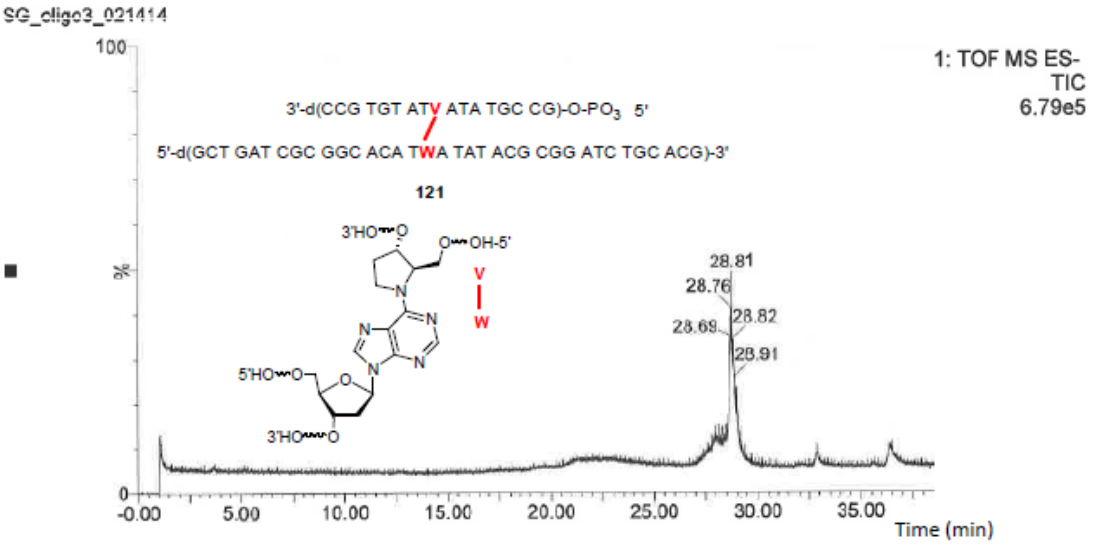
App. Figure 90. LC/MS of oligonucleotide **77**. Extracted ion chromatogram of **77** with the oligonucleotide sequence and chemical structure of cross-link are shown.



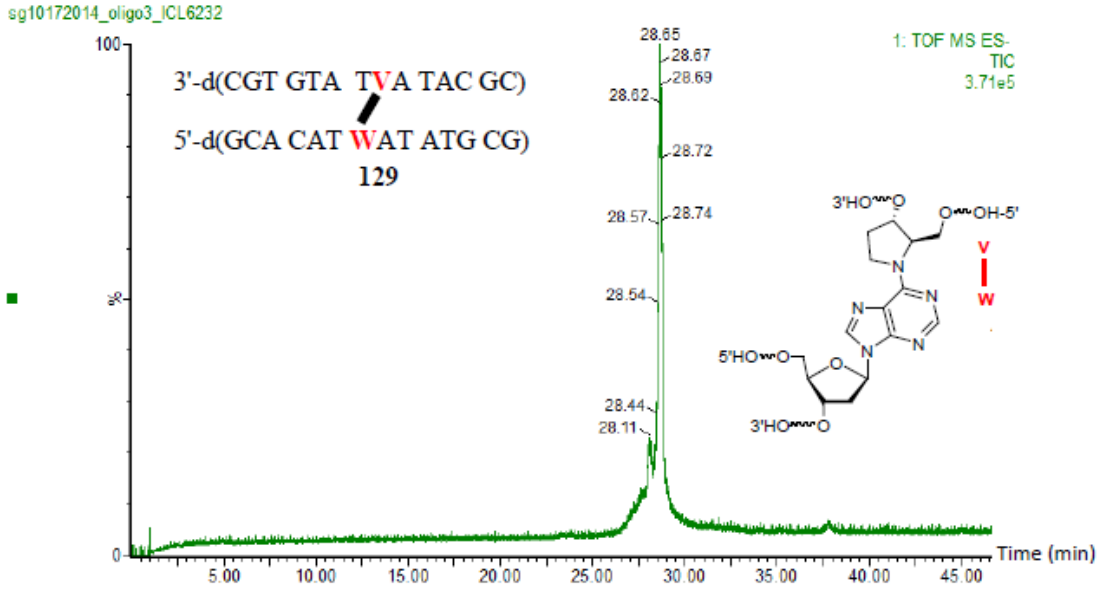
App. Figure 91. LC/MS of oligonucleotide **83**. Extracted ion chromatogram of **83** with the oligonucleotide sequence and chemical structure of cross-link are shown.



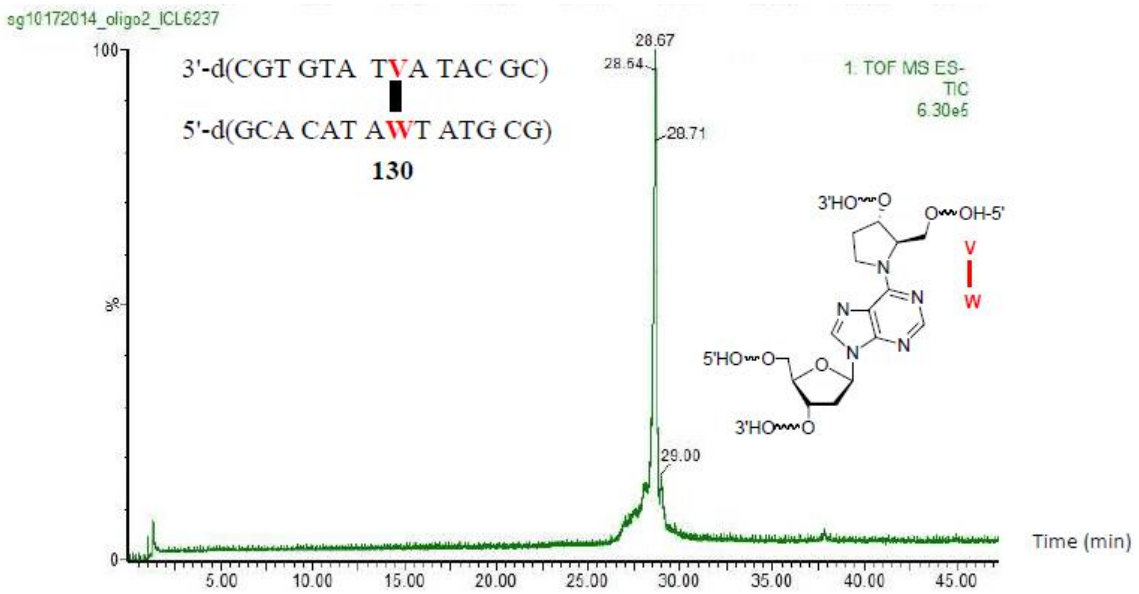
App. Figure 92. LC/MS of oligonucleotide **84**. Extracted ion chromatogram of **84** with the oligonucleotide sequence and chemical structure of cross-link are shown.



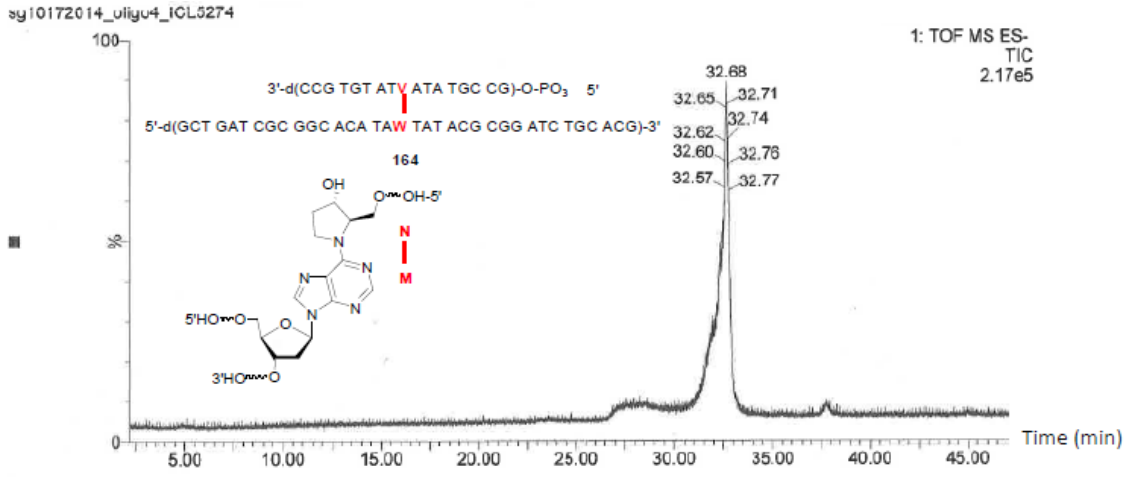
App. Figure 93. LC/MS of oligonucleotide **121**. Extracted ion chromatogram of **121** with the oligonucleotide sequence and chemical structure of cross-link are shown.



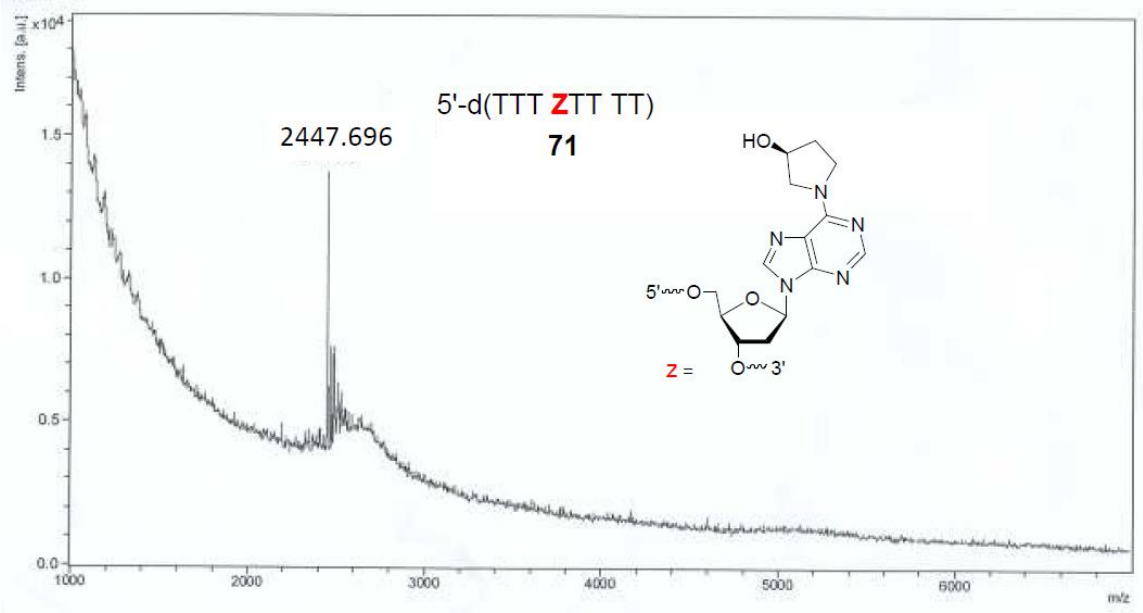
App. Figure 94. LC/MS of oligonucleotide **129**. Extracted ion chromatogram of **129** with the oligonucleotide sequence and chemical structure of cross-link are shown.



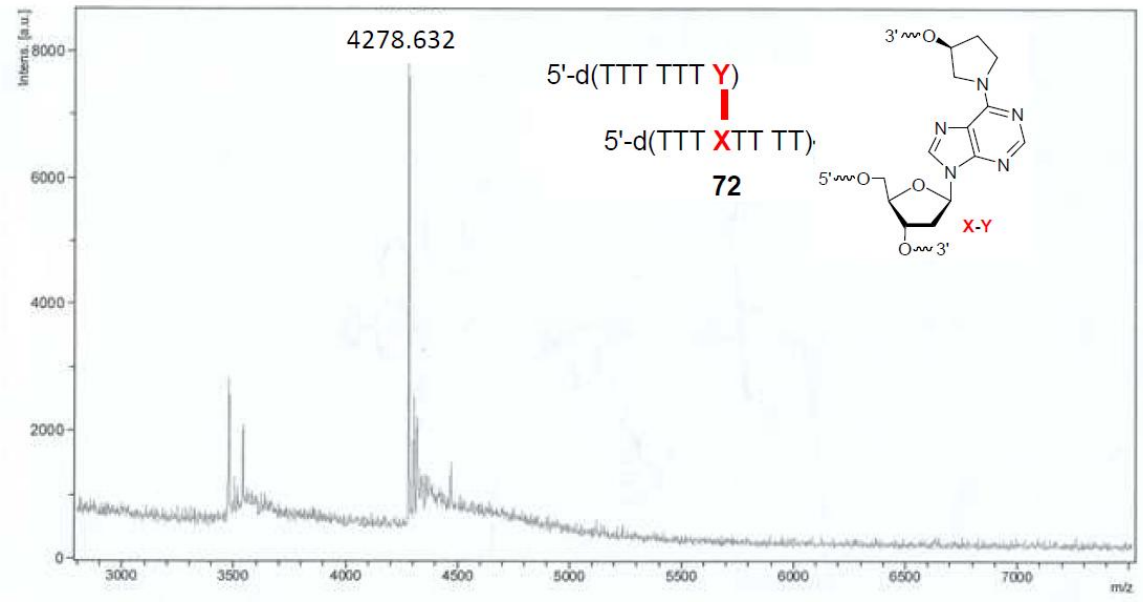
App. Figure 95. LC/MS of oligonucleotide **130**. Extracted ion chromatogram of **130** with the oligonucleotide sequence and chemical structure of cross-link are shown.



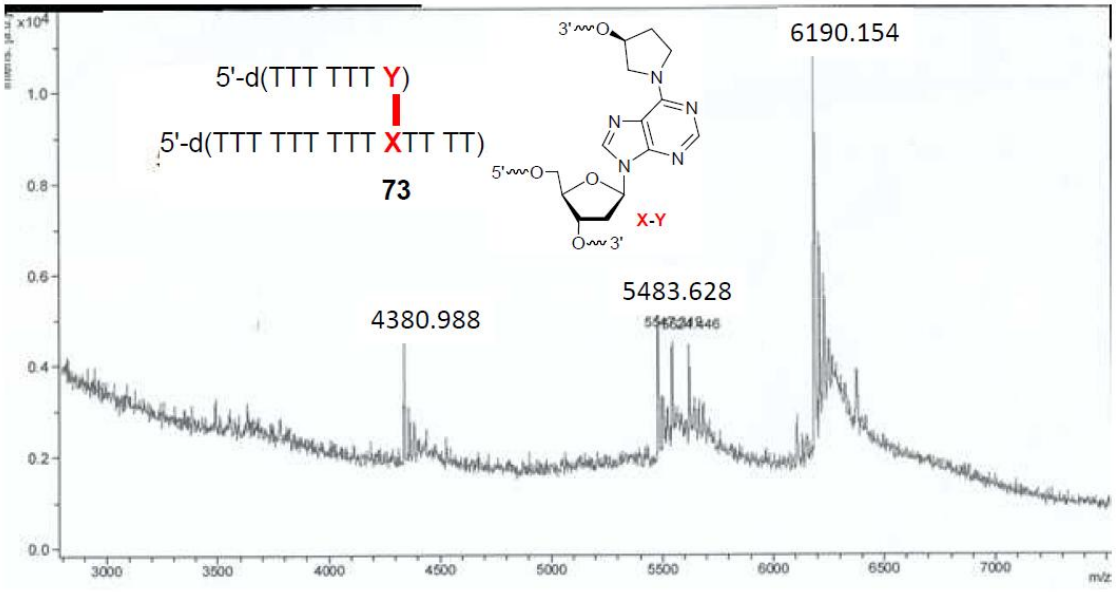
App. Figure 96. LC/MS of oligonucleotide **164**. Extracted ion chromatogram of **164** with the oligonucleotide sequence and chemical structure of cross-link are shown.



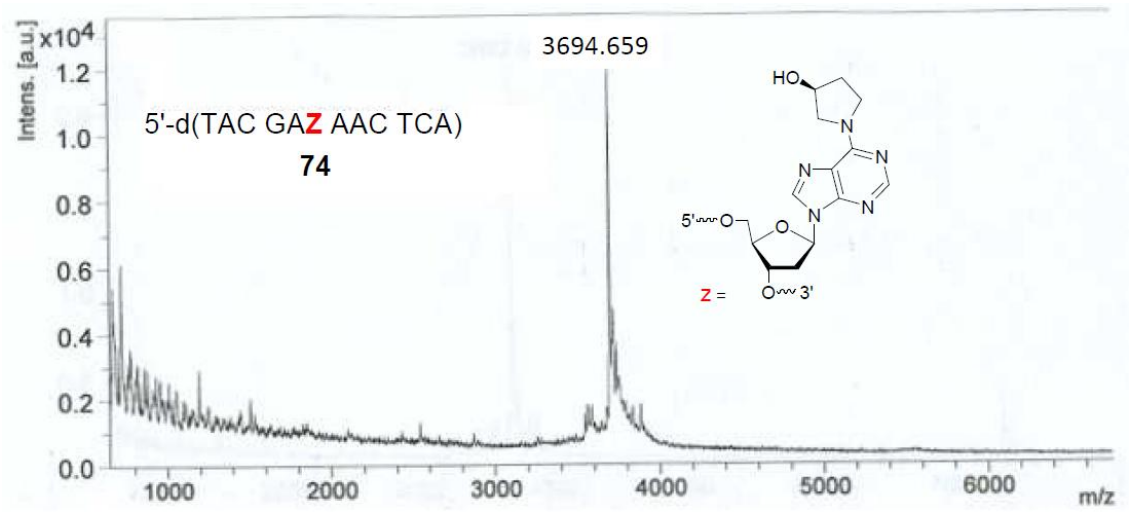
App. Figure 97. MALDI-TOF-MS of **71**. Calc'd mass 2450.622, observed mass 2447.696.



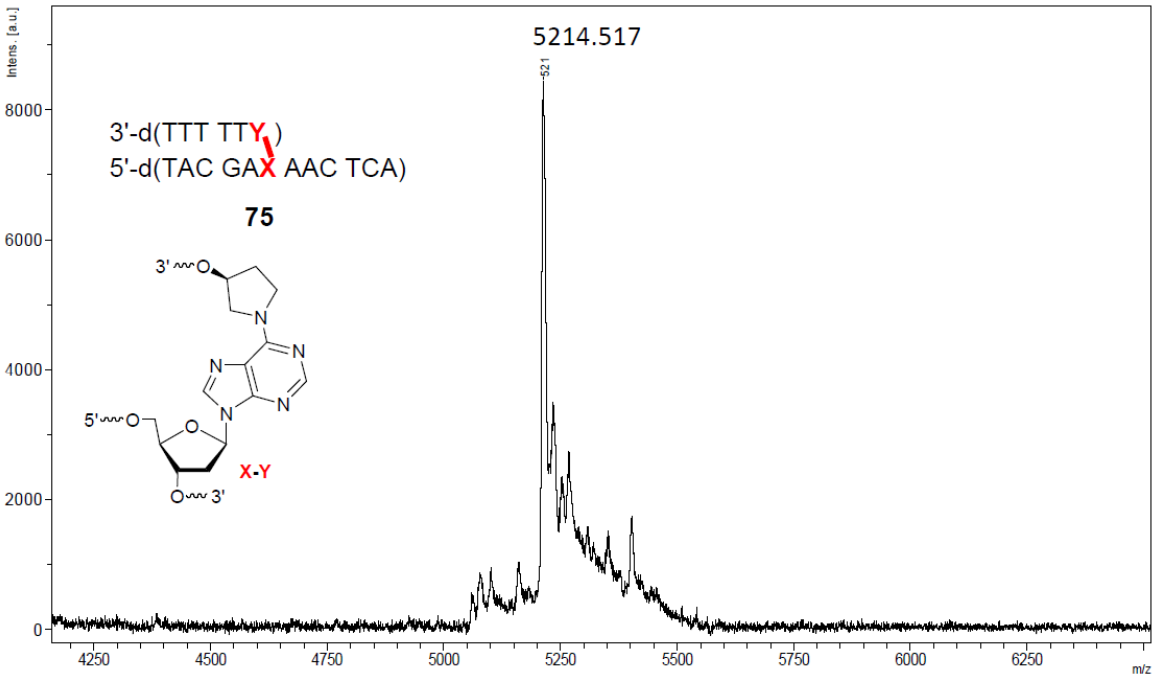
App. Figure 98. MALDI-TOF-MS of **72**. Calc'd mass 4275.806, observed mass 4278.632.



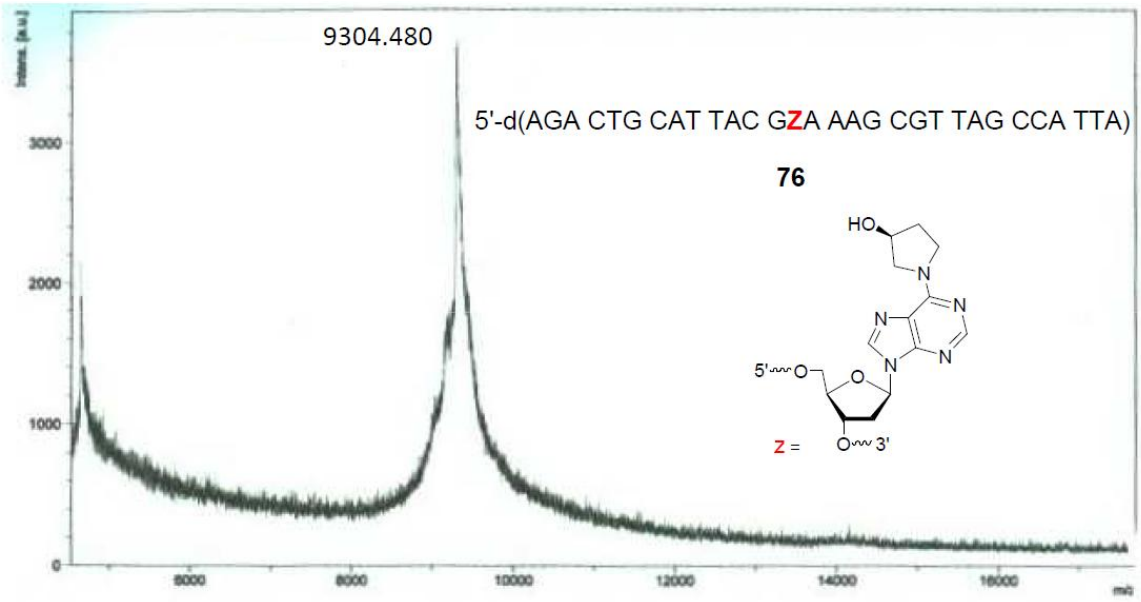
App. Figure 99. MALDI-TOF-MS of **73**. Calc'd mass 6107.621, observed mass 5483.628 [**72** + 5dT], 6190.154 [**72** + 6dT].



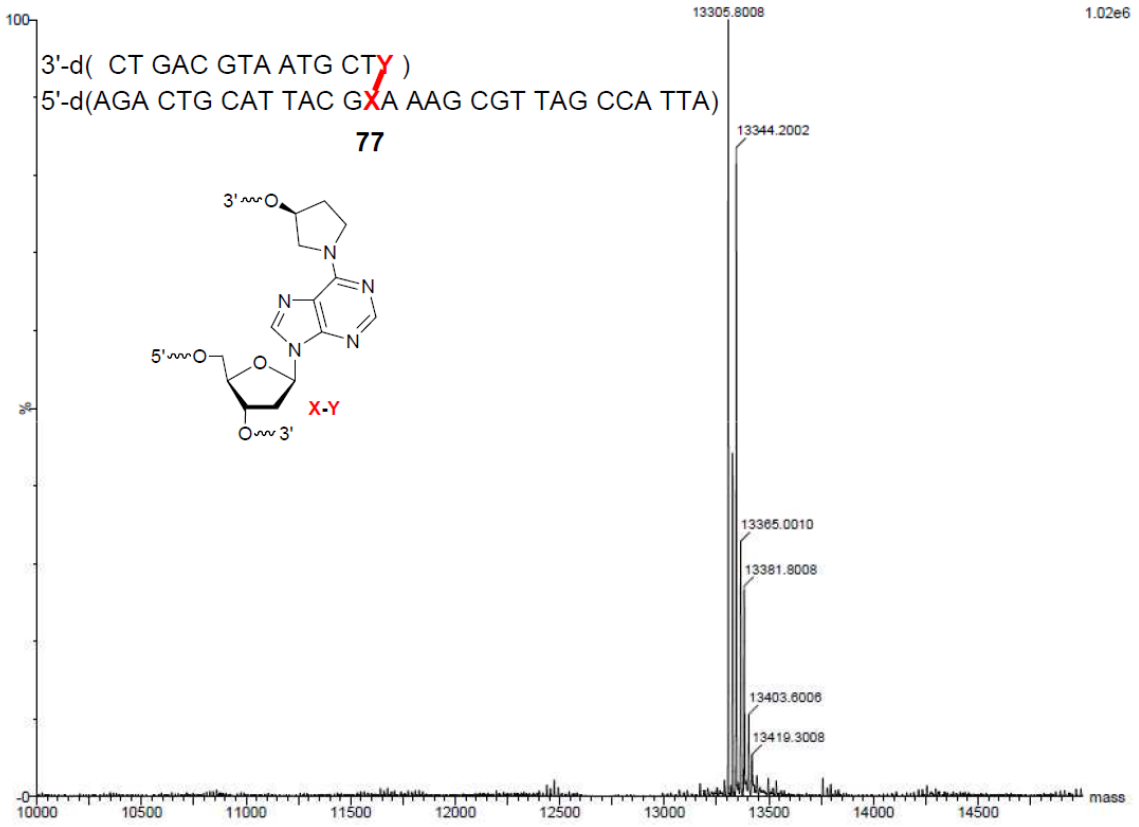
App. Figure 100. MALDI-TOF-MS of **74**. Calc'd mass 3692.454, observed mass 3694.659.



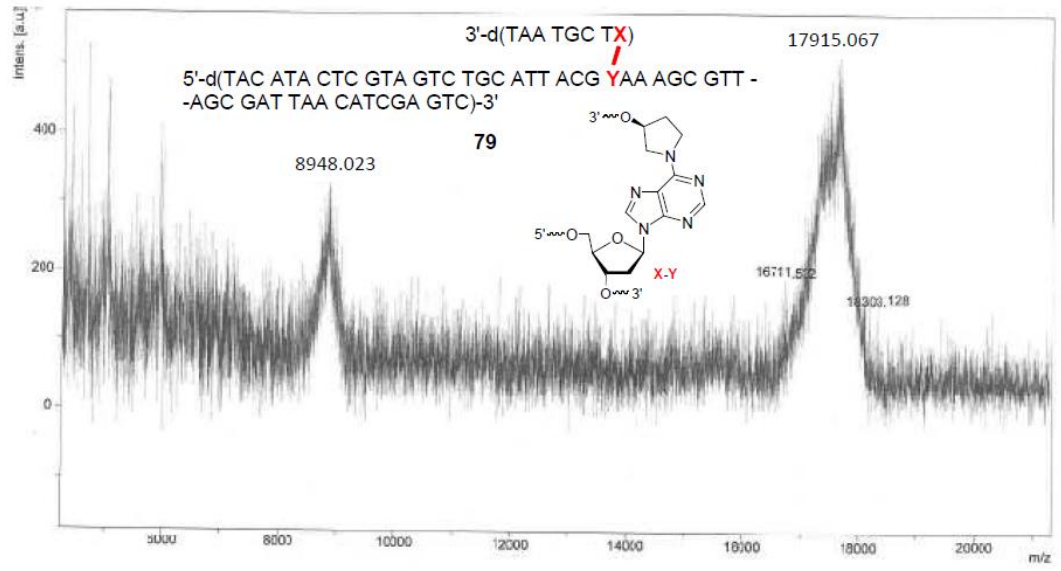
App. Figure 101. MALDI-TOF-MS of **75**. Calc'd mass 5213.422, observed mass 5214.517.



App. Figure 102. MALDI-TOF-MS of **76**. Calc'd mass 9293.094, observed mass 9304.480.

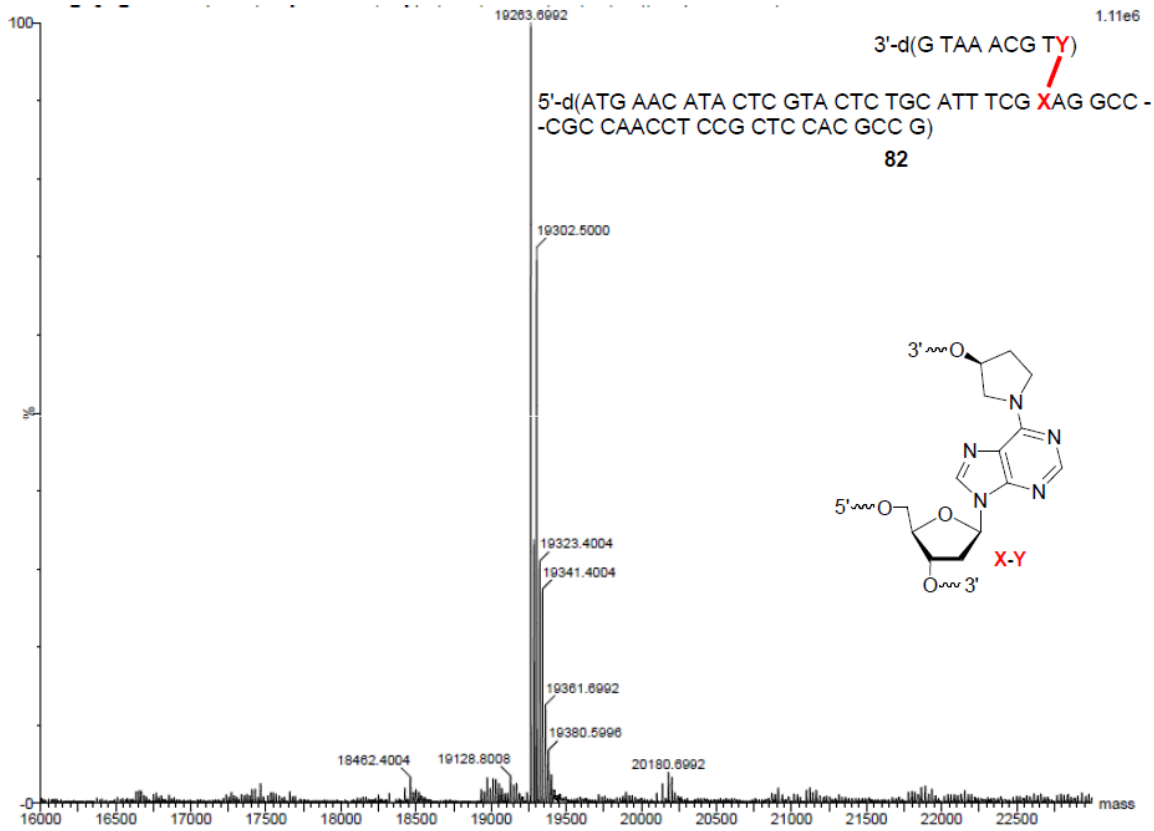


App. Figure 103. ESI-MS of **77.** Calc'd mass 13304.603, observed mass 13305.800, 13344.200 $[M + K]^+$, 13365.001 $[M + K + Na]^+$, 13381.800 $[M + 2K]^+$, 13403.600 $[M + 2K + Na]^+$, 13419.300 $[M + 3K]^+$.

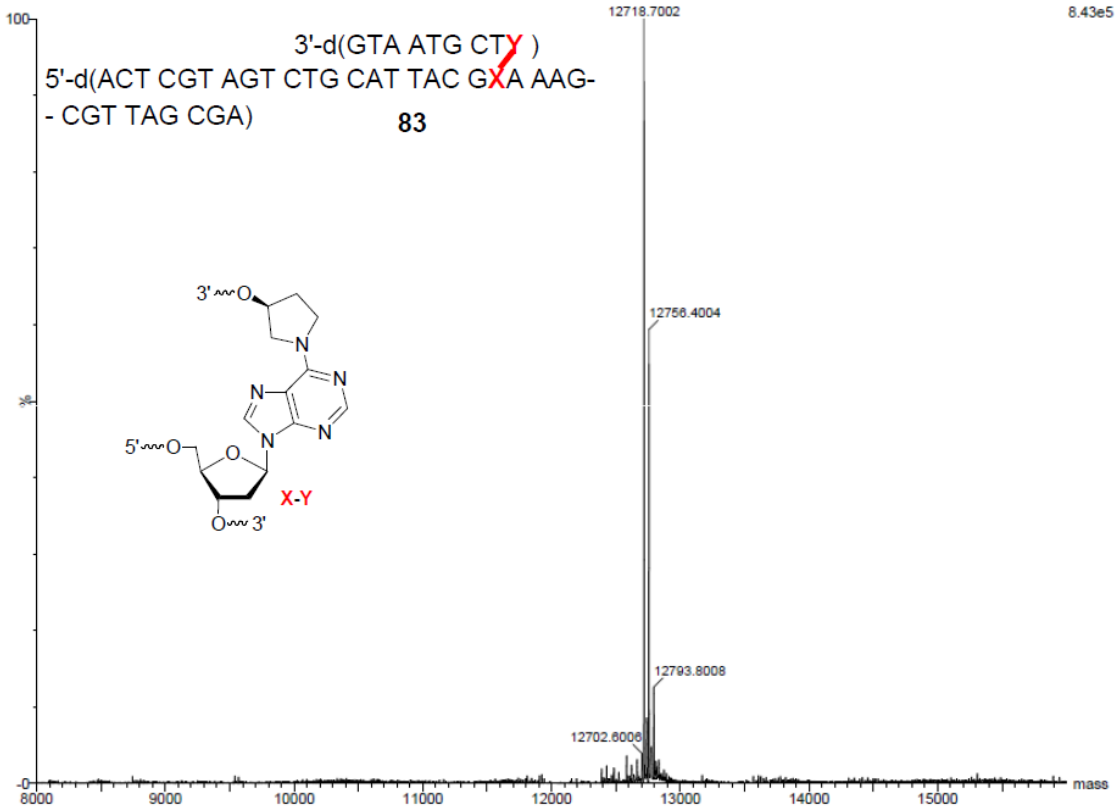


.

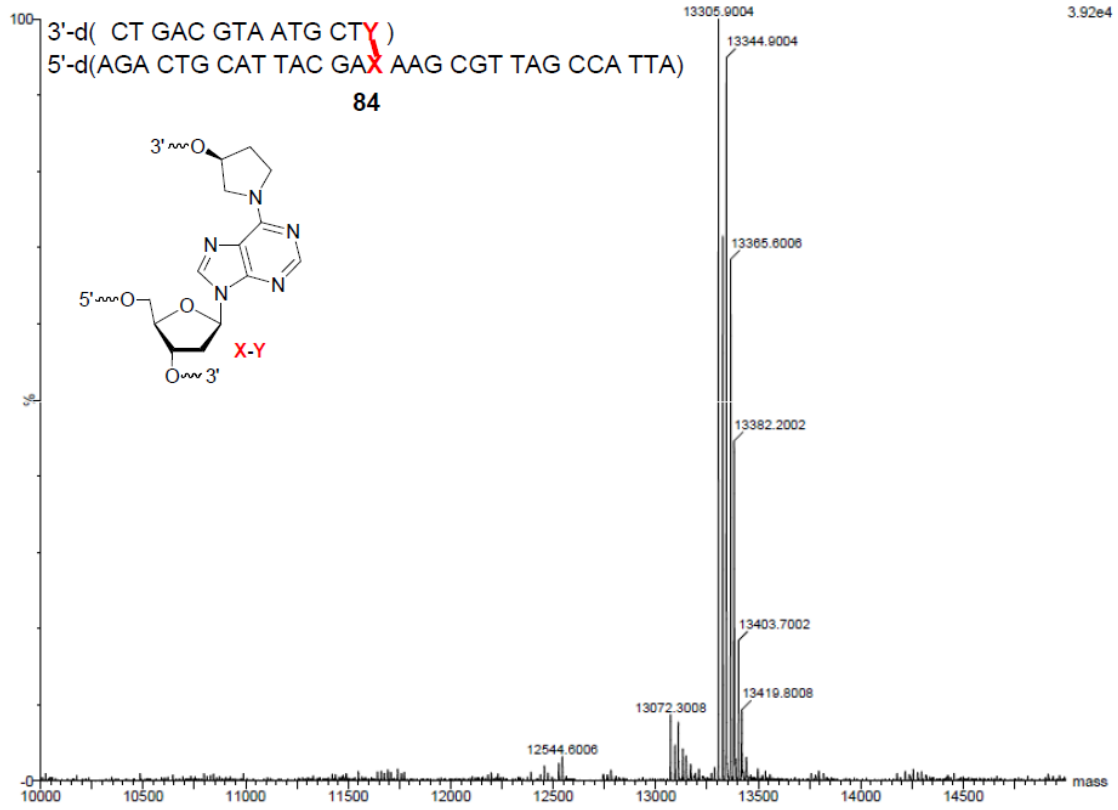
App. Figure 104. MALDI-TOF-MS of **79**. Calc'd mass 17908.693, observed mass 8948.023 (z = -2), 17915.067.



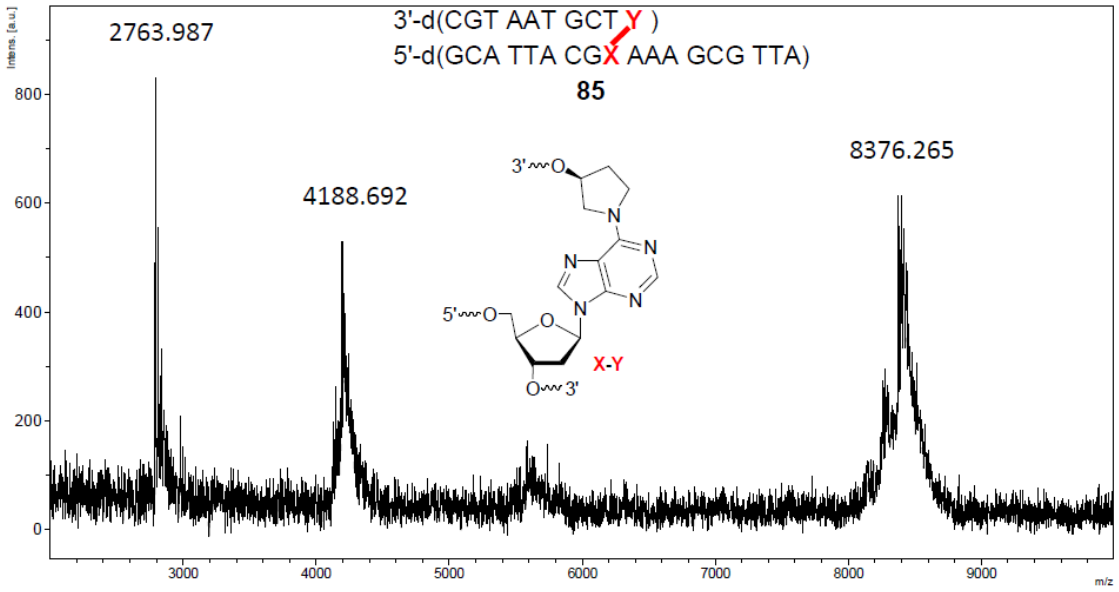
App. Figure 105. ESI-MS of **82.** Calc'd mass 19262.492, observed mass 19263.699, 19302.500 $[M + K]^+$, 19223.400 $[M + K + Na]^+$, 19341.400 $[M + 2K]^+$, 19361.699 $[M + 2K + Na]^+$, 19380.599 $[M + 3K]^+$.



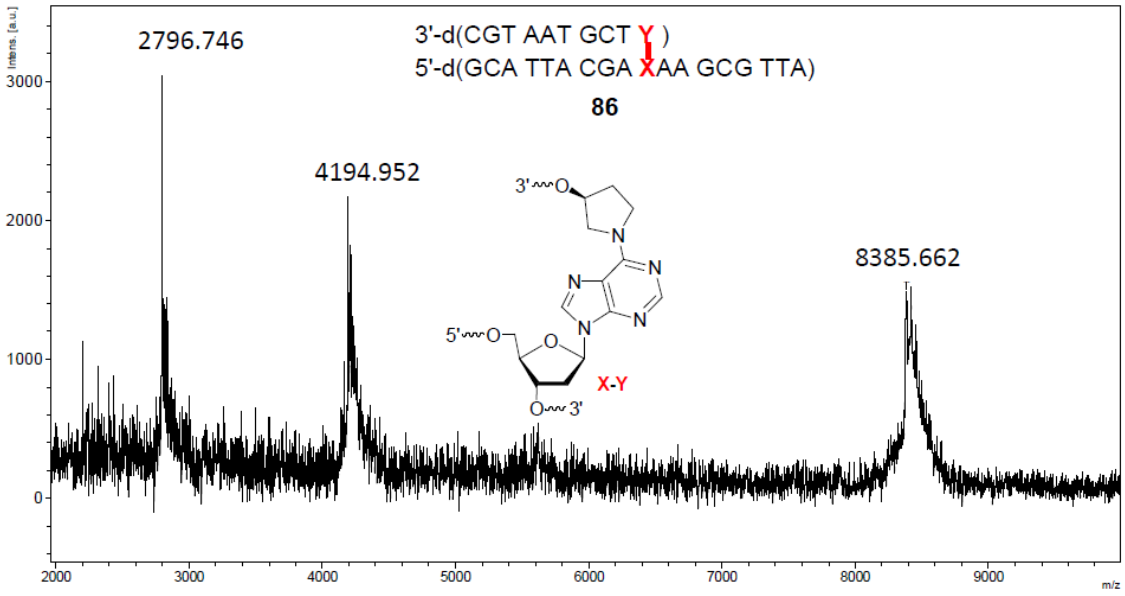
App. Figure 106. ESI-MS of **83.** Calc'd mass 12718.312, observed mass 12718.700, 12756.400 $[M + K]^+$, 12793.800 $[M + 2K]^+$.



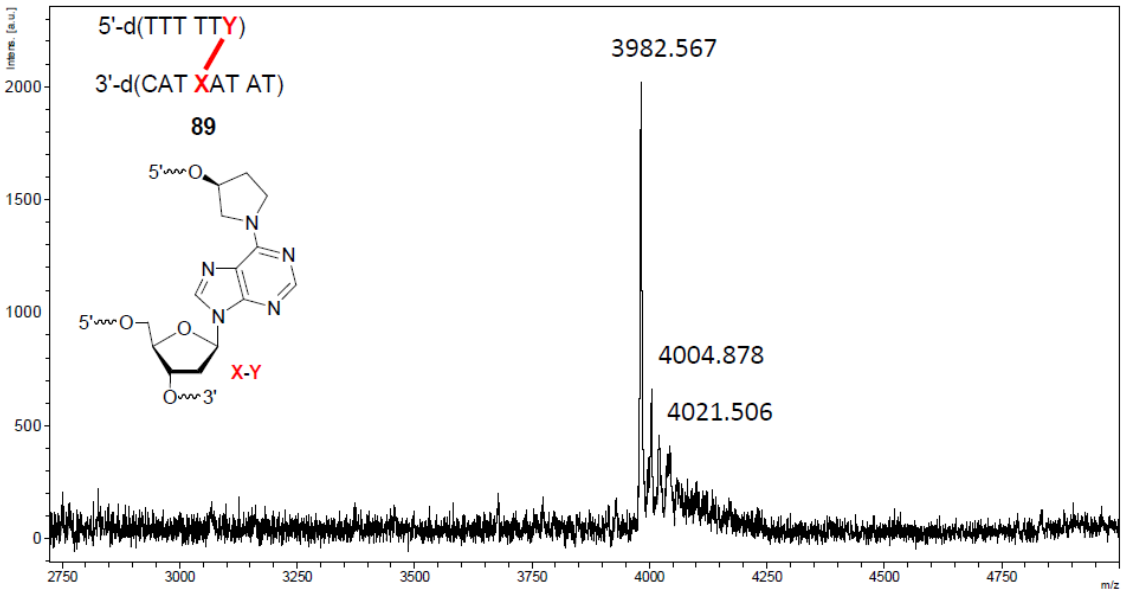
App. Figure 107. ESI-MS of **84.** Calc'd mass 13304.603, observed mass 13305.900, 13344.900 $[M + K]^+$, 13365.601 $[M + K + Na]^+$, 13382.200 $[M + 2K]^+$, 13403.700 $[M + 2K + Na]^+$, 13419.800 $[M + 3K]^+$.



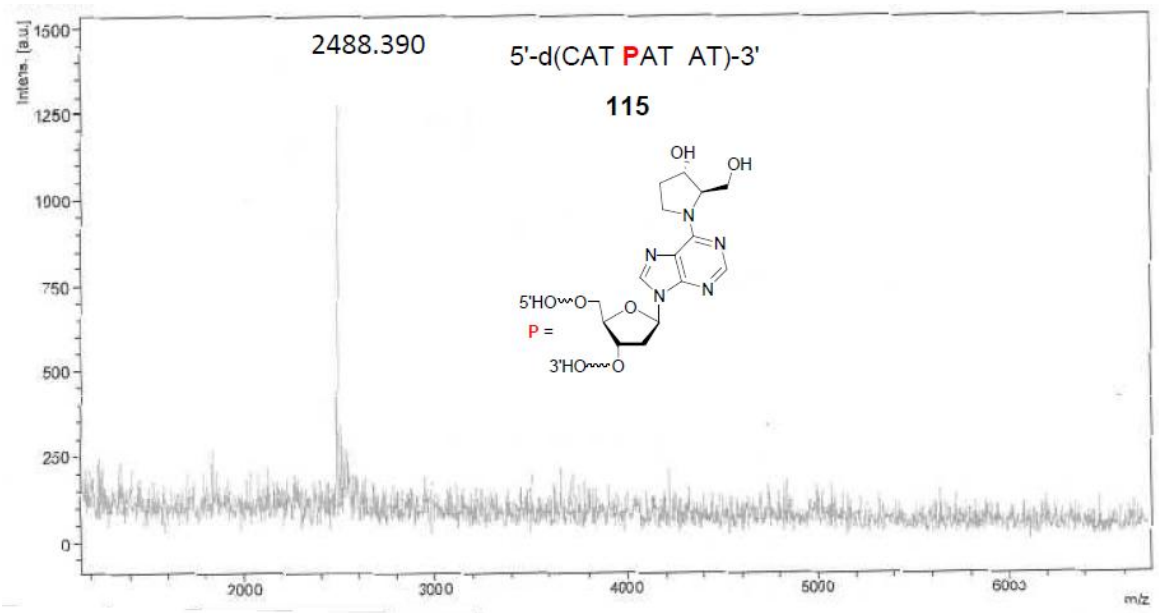
App. Figure 108. MALDI-TOF-MS of **85.** Calc'd mass 8377.514, observed mass 2763.987 (z = -3), 4188.692 (z = -2), 8376.265.



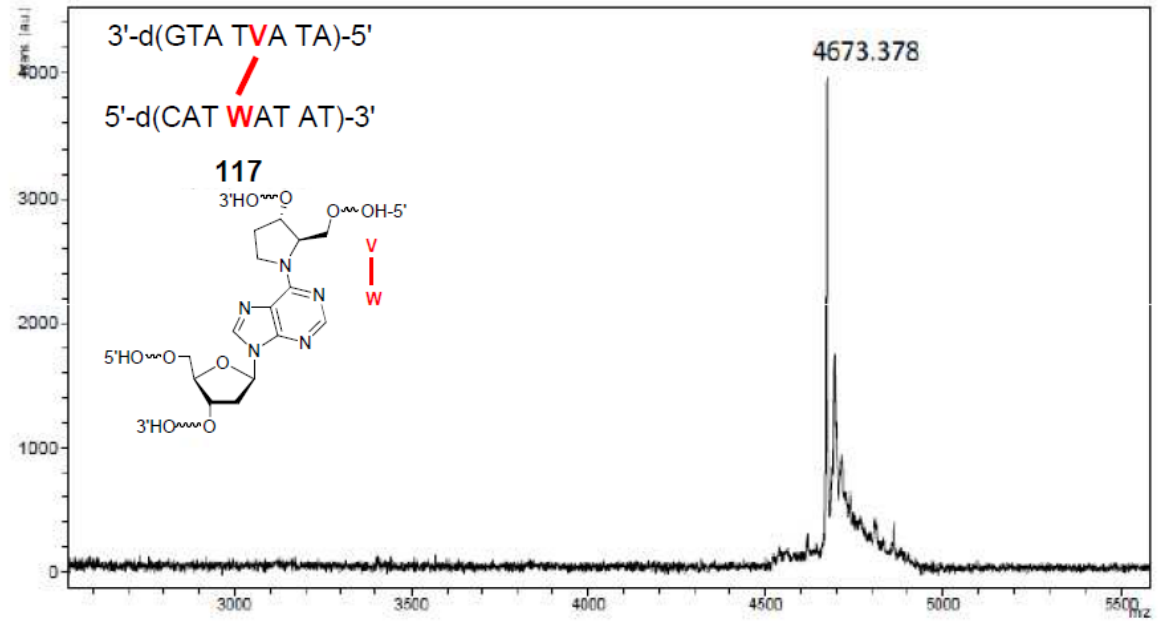
App. Figure 109. MALDI-TOF-MS of **86.** Calc'd mass 8377.514, observed mass 2796.987 (z = -3), 4194.692 (z = -2), 8385.662.



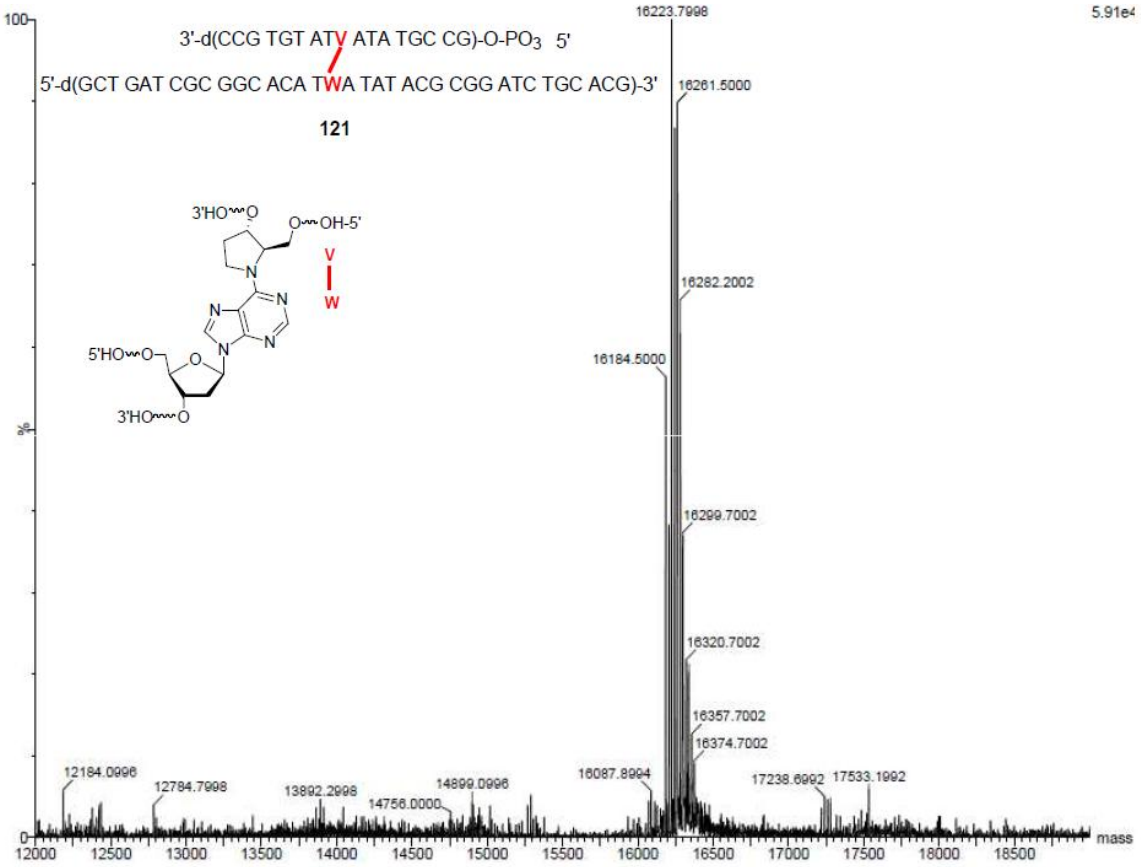
App. Figure 110. MALDI-TOF-MS of **89**. Calc'd mass 3983.611, observed mass 3982.567, 4004.878 $[M + Na]^+$, 4021.506 $[M + K]^+$.



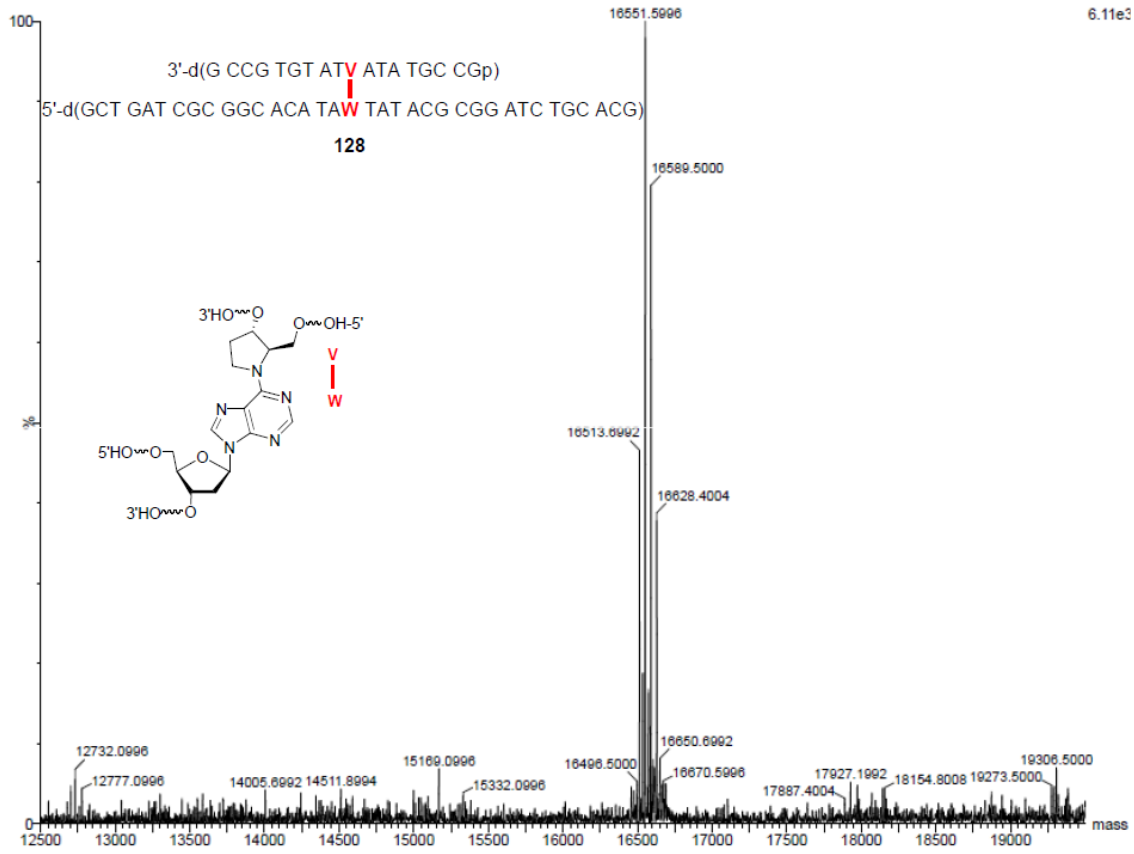
App. Figure 111. MALDI-TOF-MS of **115**. Calc'd mass 2492.651, observed mass 2488.390.



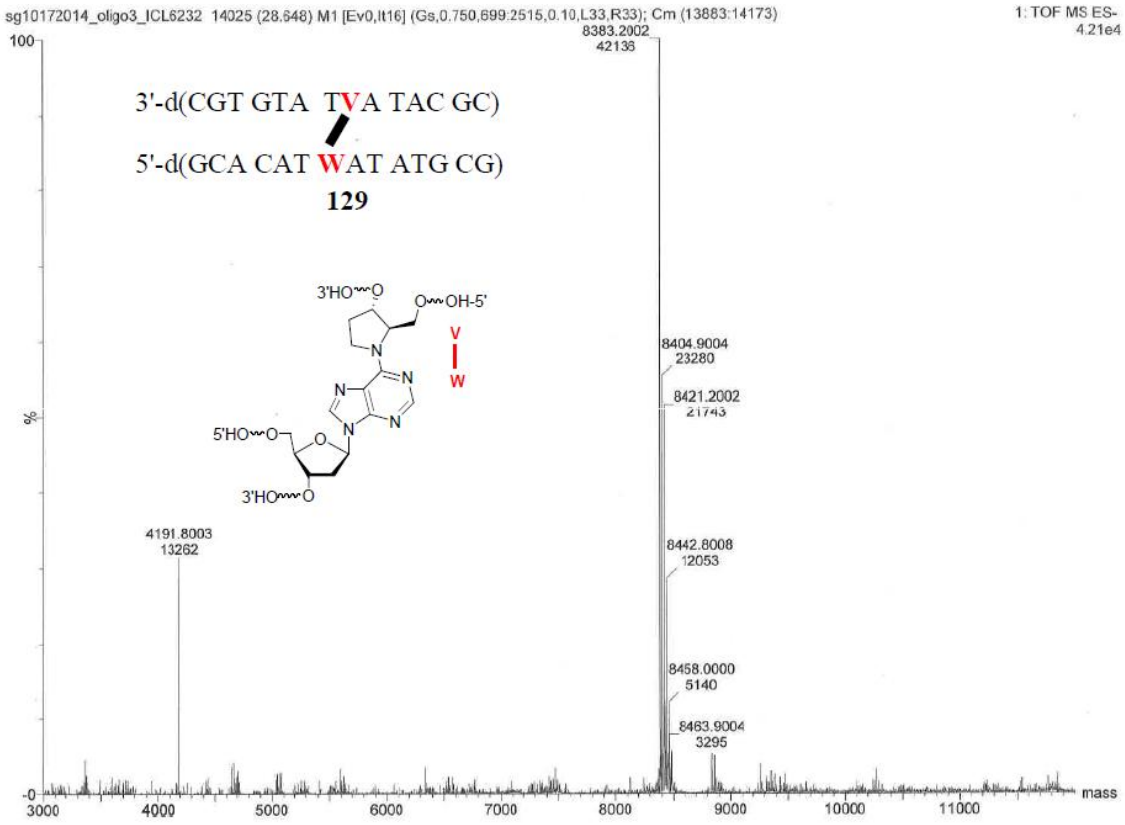
App. Figure 112. MALDI-TOF-MS of **117**. Calc'd mass 4674.102, observed mass 4673.378.



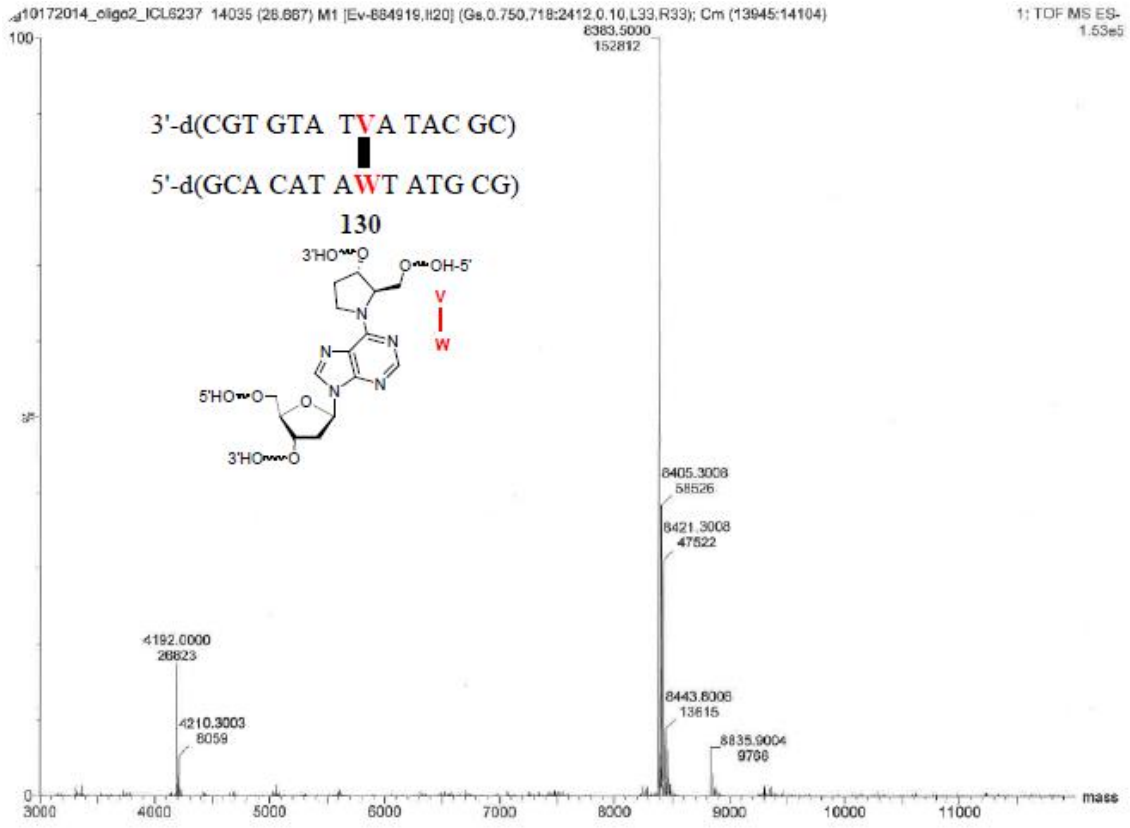
App. Figure 113. ESI-MS of **121**. Calc'd mass 16182.485, observed mass 16184.500, 16223.799 $[M + K]^+$, 16261.500 $[M + 2K]^+$, 16282.200 $[M + 2K + Na]^+$, 16299.700 $[M + 3K]^+$, 16320.700 $[M + 3K + Na]^+$, 16357.700 $[M + 4K + Na]^+$, 16374.700 $[M + 5K]^+$.



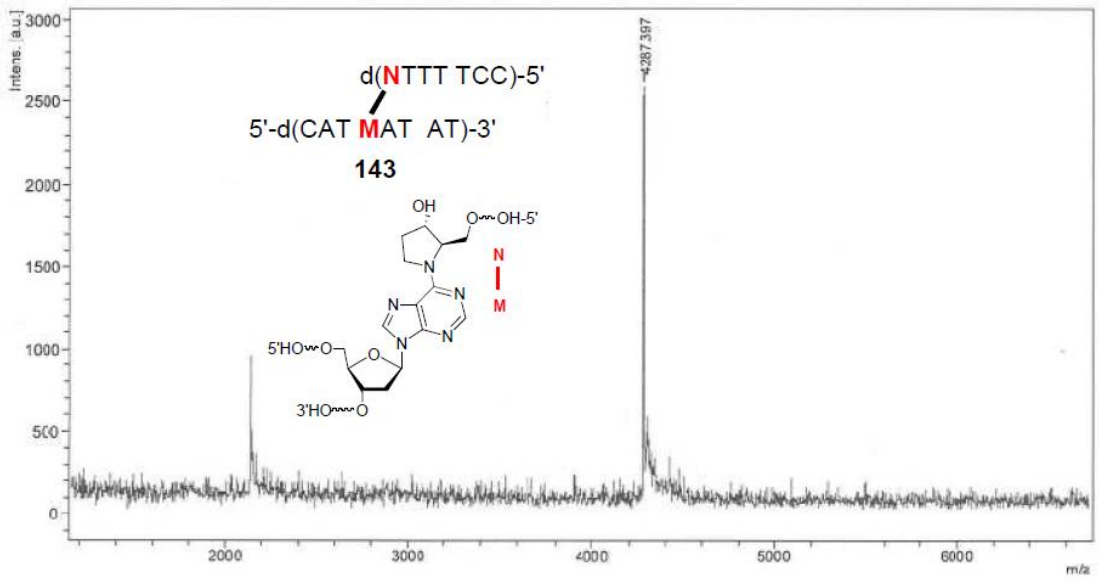
App. Figure 114. ESI-MS of **128**. Calc'd mass 16511.658, observed mass 16513.699, 16551.599 $[M + K]^+$, 16589.500 $[M + 2K]^+$, 16628.400 $[M + 3K]^+$.



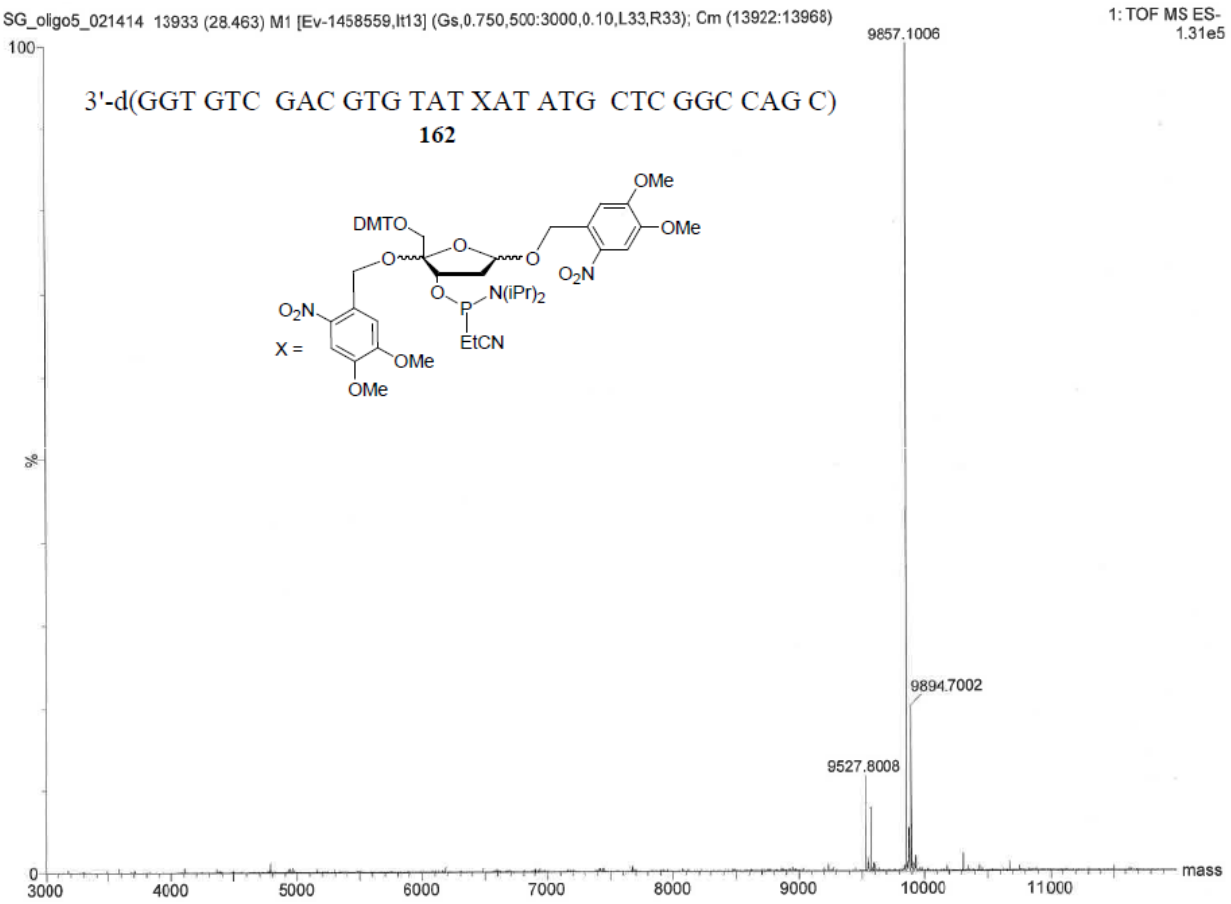
App. Figure 115. ESI-MS of **129**. Calc'd mass 8383.464, observed mass 4191.800 (z = -2), 8383.200, 8404.900 $[M + Na]^+$, 8421.200 $[M + K]^+$, 8442.800 $[M + Na + K]^+$, 8458.000 8421.200 $[M + 2K]^+$, 8463.900 $[M + 2Na + K]^+$.



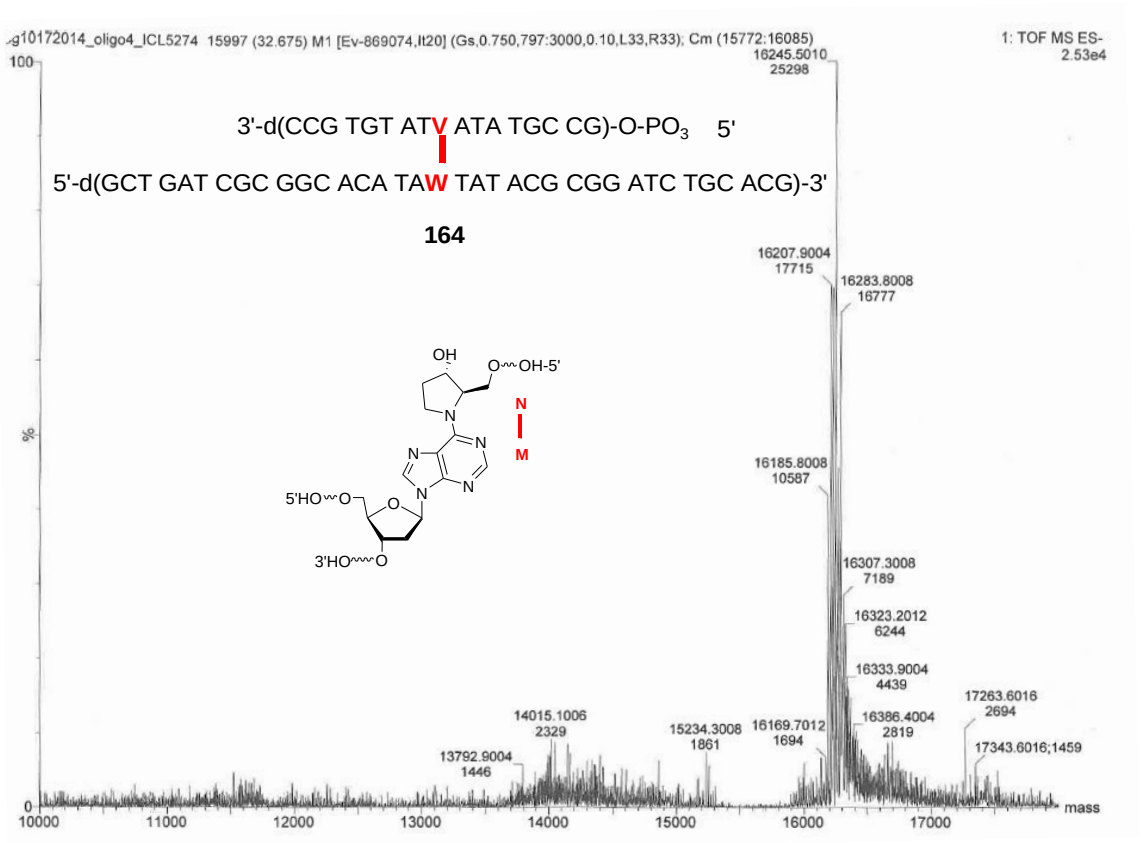
App. Figure 116. ESI-MS of **130**. Calc'd mass 8383.464, observed mass 4192.000 (z = -2), 8383.500, 8405.300 $[M + Na]^+$, 8421.300 $[M + K]^+$, 8443.800 $[M + Na + K]^+$.



App. Figure 117. MALDI-TOF-MS of **143**. Calc'd mass 4287.809, observed mass 4287.397.



App. Figure 118. ESI-MS of **162**. Calc'd mass 9855.495, observed mass 9857.100, 9894.700 $[M + K]^+$.



App. Figure 119. ESI-MS of **164**. Calc'd mass 16182.485, observed mass 16185.800, 16207.900 [M + Na]$^+$, 16245.501 [M + Na + K]$^+$, 16283.801 [M + Na + 2K]$^+$, 16307.300 [M + 2Na + 2K]$^+$, 16323.201 [M + Na + 3K]$^+$, 16333.900 [M + 3Na + 2K]$^+$, 16386.400 [M + 2Na + 4K]$^+$.

.

App. Table 1. Chemical probing experiments on DOB ICL analogue 3 for 5'-side of dA containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)	Statistical significance based on student's t test (between **155** and **156**)
A_{14} of **157** A_{15} of **157**	0.67 ± 0.04 0.79 ± 0.09	A_8 (hydroxyl radical)	The difference is significant to P value 0.003
Cross-linked position of **155** (X_{14})	0.94 ± 0.07		
Cross-linked position of **156** (X_{15})	0.75 ± 0.04		
G_{13} of **157**	0.81 ± 0.08	G_6 (DMS)	The difference is significant to P value 0.0983
G_{13} of **155**	1.25 ± 0.07		
G_{13} of **156**	1.41 ± 0.07		
C_{12} of **157**	NA	C_7 (hydroxylamine)	The difference is significant to P value 0.4839
C_{12} of **155**	0.87 ± 0.03		
C_{12} of **156**	0.93 ± 0.15		
A_{11} of **157**	1.13 ± 0.07	A_8 (DEPC)	The difference is significant to P value 0.2
A_{11} of **155**	1.03 ± 0.11		
A_{11} of **156**	1.15 ± 0.08		

App. Table 2. Chemical probing experiments on DOB ICL analogue 3 for 3'-side of dA containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference Nucleotide (reagent)	Statistical significance based on student's t test (between **155** and **156**)
A_{15} of **157** A_{16} of **157**	1.12±0.15 0.82 ± 0.07	A_{23} (DEPC)	The difference is significant to P value 0.0440
A_{15} of **158** A_{16} of **158**	4.56 ± 0.48 4.18 ± 0.62		
Position adjacent to cross-link in **155** (A_{15})	1.93 ± 0.32		
Position adjacent to cross-link in **156** (A_{16})	3.09 ± 0.85		
A_{16} of **155**	1.90 ± 0.36	A_{23} (DEPC)	The difference is significant to P value 0.0425
A_{16} of **156**	3.09 ± 0.85		
A_{17} of **157**	0.52±0.03	A_{23} (DEPC)	The difference is significant to P value 0.3169
A_{17} of **158**	1.83 ± 0.12		
A_{17} of **155**	1.17 ± 0.25		
A_{17} of **156**	1.36 ± 0.23		

App. Table 3. Chemical probing experiments on DOB ICL analogue 3 for 3'-side of abasic side analog containing strand ("top" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)	Statistical significance based on student's t test (between **155** and **156**)
T_{46} of **157**	0.41 ± 0.39	G_{54} (hydroxyl radical)	The difference is significant to P value 0.000
Y_{46} of **155**	1.87 ± 0.43		
Y_{46} of **156**	3.77 ± 0.25		
T_{47} of **157**	0.81 ± 0.11	G_{54} ($KMnO_4$)	The difference is significant to P value 0.000
T_{47} of **155**	18.31 ± 1.23		
T_{47} of **156**	6.22 ± 0.62		
C_{48} of **157**	0.76 ± 0.22	C_{55} (hydroxyl amine)	The difference is significant to P value 0.136
C_{48} of **155**	0.89 ± 0.02		
C_{48} of **156**	1.15 ± 0.30		
G_{49} of **157**	1.56 ± 0.02	G_{54} (DMS)	The difference is significant to P value 0.008
G_{49} of **155**	1.45 ± 0.22		
G_{49} of **156**	1.90 ± 0.04		
T_{50} of **157**	1.84 ± 0.12	G_{54} ($KMnO_4$)	The difference is significant to P value 0.002
T_{50} of **155**	1.48 ± 0.08		
T_{50} of **156**	2.14 ± 0.27		

App. Table 4. Chemical probing experiments on DOB ICL analogue 3 for 5'-side of abasic side analog containing strand ("top" strand)

Nucleotide of interest in observed ICL (**155**), unobserved ICL (**156**) and control duplex (**157**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)	Statistical significance based on student's t test (between **155** and **156**)
T_{45} of **157**	0.72 ± 0.23	T_{38} ($KMnO_4$)	The difference is significant to P value 0.0842
T_{45} of **158**	21.23 ± 0.89		
T_{45} of **155**	6.25 ± 0.44		
T_{45} of **156**	7.20 ± 0.56		
T_{44} of **157**	0.65 ± 0.23	T_{38} ($KMnO_4$)	The difference is significant to P value 0.001
T_{44} of **158**	9.59 ± 0.63		
T_{44} of **155**	2.59 ± 0.17		
T_{44} of **156**	5.12 ± 0.22		

App. Table 5. Chemical probing experiments on C4-AP ICL analogue 4 for 5'-side of dA containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**135**), unobserved ICL (**170**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)	Statistical significance based on student's t test (between **135** and **170**)
T_{16} of **167**	0.84 ± 0.05	T_{18} (T_6 for **167**) ($KMnO_4$)	The difference is significant to P value 0.008
T_{28} of **135**	1.42 ± 0.02		
T_{28} of **170**	1.60 ± 0.09		
A_{15} of **167**	0.77 ± 0.17	A_{17} (A_5 for **167**) (DEPC)	The difference is significant to P value 0.001
A_{27} of **135**	2.43 ± 0.03		
A_{27} of **170**	2.54 ± 0.02		

App. Table 6. Chemical probing experiments on C4-AP ICL analogue 4 for 3'-side of dA containing strand ("bottom" strand)

Nucleotide of interest in observed ICL (**165**), unobserved ICL (**166**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (reagent)	Statistical significance based on student's t test (between **165** and **166**)
A_{17} of **167**	1.12 ± 0.33	A_{28} (hydroxyl radical cleavage	The difference is significant to P value 0.13
A_{18} of **167**	1.04 ± 0.23		
W_{17} of **165**	0.67 ± 0.11		
W_{18} of **166**	0.94 ± 0.39		
T_{19} of **167**	1.35 ± 0.17	T_{29} ($KMnO_4$)	The difference is significant to P value 0.000
T_{19} of **165**	14.99 ± 0.84		
T_{19} of **166**	2.61 ± 0.27		
A_{20} of **167**	1.47 ± 0.26	A_{28} (DEPC)	The difference is significant to P value 0.002
A_{20} of **165**	3.39 ± 0.21		
A_{20} of **166**	4.80 ± 0.47		
A_{18} of **167**	Not calculated	A_{28} (DEPC)	The difference is significant to P value 0.02
A_{20} of **167**	1.47 ± 0.26		
A_{18} of **165**	4.01 ± 0.19		
A_{20} of **166**	4.80 ± 0.47		
T_{21} of **167**	0.97 ± 0.24	T_{29} ($KMnO_4$)	The difference is significant to P value 0.59
T_{21} of **165**	1.85 ± 0.14		
T_{21} of **166**	1.80 ± 0.11		

App. Table 7. Chemical probing experiments on C4-AP ICL analogue 4 for 5'-side of abasic site analogue containing strand ("top" strand)

Nucleotide of interest in observed ICL (**165**), unobserved ICL (**166**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide (Reagent)	Statistical significance based on student's t test (between **165** and **166**)
A_{54} of **167**	3.01± 0.34	G_{48} (DEPC)	The difference is significant to P value 0.000
A_{54} of **165**	5.47 ± 0.35		
A_{54} of **166**	18.91 ± 6.67		
T_{53} of **167**	0.94 ± 0.03	T_{45} ($KMnO_4$)	The difference is significant to P value 0.000
T_{53} of **165**	2.39 ± 0.27		
T_{53} of **166**	3.85 ± 0.17		
A_{52} of **167**	1.73 ± 0.37	G_{48} (DEPC)	The difference is significant to P value 0.000
A_{52} of **165**	2.44 ± 0.33		
A_{52} of **166**	6.67 ± 1.68		

App. Table 8. Chemical probing experiments on C4-AP ICL analogue 4 for 3'-side of abasic site analogue containing strand ("top" strand)

Nucleotide of interest in observed ICL (**165**), unobserved ICL (**166**) and control duplex (**167**)	Normalized cleavage (NC) value [= $\frac{\text{(Cleavage band volume of nucleotide of interest)}}{\text{(Cleavage band volume of reference nucleotide)}}$]	Reference nucleotide	Statistical significance based on student's t test (between **165** and **166**)
A_{55} of **167**	1.00 ± 0.21	A_{67} (hydroxyl radical cleavage)	The difference is significant to P value 0.67
V_{55} of **165**	0.84 ± 0.38		
V_{55} of **166**	0.77 ± 0.10		
T_{56} of **167**	2.31 ± 0.22	T_{68} ($KMnO_4$)	The difference is significant to P value < 0.00001
T_{56} of **165**	1.36 ± 0.12		
T_{56} of **166**	4.45 ± 0.30		
A_{57} of **167**	0.95 ± 0.24	A_{67} (DEPC)	The difference is significant to P value 0.0004
A_{57} of **165**	2.08 ± 0.15		
A_{57} of **166**	2.72 ± 0.09		
T_{58} of **167**	1.28 ± 0.23	T_{68} ($KMnO_4$)	The difference is significant to P value 0.0002
T_{58} of **165**	1.07 ± 0.13		
T_{58} of **166**	1.70 ± 0.10		

7. Bibliography

(1) Rink, S. M.; Hopkins, P. B. *Biochemistry* **1995**, *34*, 1439–1445.

(2) Wiencke, J. K.; Wiemels, J. *Mutat. Res. Genet. Toxicol.* **1995**, *339*, 91–119.

(3) Suresh Kumar, G.; Lipman, R.; Cummings, J.; Tomasz, M. *Biochemistry* **1997**, *36*, 14128–14136.

(4) Dall'Aqua, F.; Marciani, S.; Vedaldi, D.; Rodighiero, G. *FEBS Lett.* **1972**, *27*, 192–194.

(5) Mukai, F. H.; Goldstein, B. D. *Science* **1976**, *191*, 868–869.

(6) Johnson, K. M.; Price, N. E.; Wang, J.; Fekry, M. I.; Dutta, S.; Seiner, D. R.; Wang, Y.; Gates, K. S. *J. Am. Chem. Soc.* **2013**, *135*, 1015–1025.

(7) Price, N. E.; Johnson, K. M.; Wang, J.; Fekry, M. I.; Wang, Y.; Gates, K. S. *J. Am. Chem. Soc.* **2014**, *136*, 3483–3490.

(8) Sczepanski, J. T.; Jacobs, A. C.; Greenberg, M. M. *J. Am. Chem. Soc.* **2008**, *130*, 9646–9647.

(9) Guan, L.; Greenberg, M. M. *J. Am. Chem. Soc.* **2009**, *131*, 15225–15231.

(10) Bessho, T. *J. Biol. Chem.* **2003**, *278*, 5250–5254.

(11) Ahn, B.; Kang, D.; Kim, H.; Wei, Q. *Mol. Cells* **2004**, *18*, 249–255.

(12) Sasaki, M. S.; Tonomura, A. *Cancer Res.* **1973**, *33*, 1829–1836.

(13) Wang, X.; Peterson, C. A.; Zheng, H.; Nairn, R. S.; Legerski, R. J.; Li, L. *Mol. Cell. Biol.* **2001**, *21*, 713–720.

(14) Cassier, C.; Chanet, R.; Moustacchi, E. *Photochem. Photobiol.* **1984**, *39*, 799–803.

(15) Gilman, A.; Philips, F. S. *Science* **1946**, *103*, 409–436.

(16) Hofheinz, R.-D.; Beyer, U.; Al-Batran, S.-E.; Hartmann, J. T. *Onkologie* **2008**, *31*, 271–281.

(17) Stern, R. S. *N. Engl. J. Med.* **2007**, *357*, 682–690.

(18) Noll, D. M.; Mason, T. M.; Miller, P. S. *Chem. Rev.* **2006**, *106*, 277–301.

(19) Pu, W.; Kahn, R.; Munn, M.; Rupp, W. *J. Biol. Chem.* **1989**, *264*, 20697–20704.

(20) Weng, M.; Zheng, Y.; Jasti, V. P.; Champeil, E.; Tomasz, M.; Wang, Y.; Basu, A. K.; Tang, M. *Nucleic Acids Res.* **2010**, *38*, 6976–6984.

(21) Peng, X.; Ghosh, A. K.; Van Houten, B.; Greenberg, M. M. *Biochemistry* **2010**, *49*, 11–19.

(22) Sczepanski, J. T.; Jacobs, A. C.; Van Houten, B.; Greenberg, M. M. *Biochemistry* **2009**, *48*, 7565–7567.

(23) Dizdaroglu, M.; Schulte-Frohlinde, D.; von Sonntag, C. *Z. Naturforsch. C.* **1977**, *32*, 1021–1022.

(24) Henner, W. D.; Grunberg, S. M.; Haseltine, W. A. *J. Biol. Chem.* **1982**, *257*, 11750–11754.

(25) Dedon, P. C.; Goldberg, I. H. *Chem. Res. Toxicol.* **1992**, *5*, 311–332.

(26) Kallama, S.; Hemminki, K. *Chem. Biol. Interact.* **1986**, *57*, 85–96.

(27) Fischhaber, P. L.; Gall, A. S.; Duncan, J. A.; Hopkins, P. B. *Cancer Res.* **1999**, *59*, 4363–4368.

(28) Stone, M. P.; Cho, Y.-J.; Huang, H.; Kim, H.-Y.; Kozekov, I. D.; Kozekova, A.; Wang, H.; Minko, I. G.; Lloyd, R. S.; Harris, T. M.; Rizzo, C. J. *Acc. Chem. Res.* **2008**, *41*, 793–804.

(29) Mukherjee, S.; Guainazzi, A.; Schärer, O. D. *Nucleic Acids Res.* **2014**, *42*, 7429–7435.

(30) Op de Beeck, M.; Madder, A. *J. Am. Chem. Soc.* **2011**, *133*, 796–807.

(31) Sczepanski, J. T.; Jacobs, A. C.; Majumdar, A.; Greenberg, M. M. *J. Am. Chem. Soc.* **2009**, *131*, 11132–11139.

(32) Noronha, A. M.; Noll, D. M.; Wilds, C. J.; Miller, P. S. *Biochemistry* **2002**, *41*, 760–771.

(33) Sun, G.; Noronha, A.; Wilds, C. *Tetrahedron* **2012**, *68*, 7787–7793.

(34) Kaldor, J. M.; Day, N. E.; Hemminki, K. *Eur. J. Cancer Clin. Oncol.* **1988**, *24*, 703–711.

(35) Balcome, S.; Park, S.; Quirk Dorr, D. R.; Hafner, L.; Phillips, L.; Tretyakova, N. *Chem. Res. Toxicol.* **2004**, *17*, 950–962.

(36) Mehta, J. R.; Przybylski, M.; Ludlum, D. B. *Cancer Res.* **1980**, *40*, 4183–4186.

(37) Johannessen, T.-C. A.; Bjerkvig, R.; Tysnes, B. B. *Cancer Treat. Rev.* **2008**, *34*, 558–567.

(38) Tomasz, M.; Lipman, R.; Chowdary, D.; Pawlak, J.; Verdine, G.; Nakanishi, K. *Science (80-.).* **1987**, *235*, 1204–1208.

(39) Tomasz, M.; Chowdary, D.; Lipman, R.; Shimotakahara, S.; Veiro, D.; Walker, V.; Verdine, G. L. *Proc. Natl. Acad. Sci.* **1986**, *83*, 6702–6706.

(40) Warren, A. J.; Hamilton, J. W. *Chem. Res. Toxicol.* **1996**, *9*, 1063–1071.

(41) Borowy-Borowski, H.; Lipman, R.; Chowdary, D.; Tomasz, M. *Biochemistry* **1990**, *29*, 2992–2999.

(42) Mohammadparast, B.; Rustaiee, A. R.; Rasouli, M.; Zardari, S.; Agrawal, V. *Pharm. Biol.* **2014**, 1–4.

(43) Hata, T.; Hoshi, T.; Kanamori, K.; Matsumae, A.; Sano, Y.; Shima, T.; Sugawara, R. *J. Antibiot. (Tokyo).* **1956**, *9*, 141–146.

(44) Cimino, G. D.; Gamper, H. B.; Isaacs, S. T.; Hearst, J. E. *Annu. Rev. Biochem.* **1985**, *54*, 1151–1193.

(45) Spielmann, H. P.; Sastry, S. S.; Hearst, J. E. *Proc. Natl. Acad. Sci.* **1992**, *89*, 4514–4518.

(46) Nair, U.; Bartsch, H.; Nair, J. *Free Radic. Biol. Med.* **2007**, *43*, 1109–1120.

(47) Hoffmann, D.; Wynder, E. L. *IARC Sci. Publ.* **1986**, 145–165.

(48) Esterbauer, H. *Am J Clin Nutr* **1993**, *57*, 779S – 785.

(49) Smith, R. A.; Cohen, S. M.; Lawson, T. A. *Carcinogenesis* **1990**, *11*, 497–498.

(50) J. Marnett, L.; Hurd, H. K.; Hollstein, M. C.; Levin, D. E.; Esterbauer, H.; Ames, B. N. *Mutat. Res. Mol. Mech. Mutagen.* **1985**, *148*, 25–34.

(51) Basu, A. K.; Marnett, L. J. *Carcinogenesis* **1983**, *4*, 331–333.

(52) Yau, T. M. *Mech. Ageing Dev.* **1979**, *11*, 137–144.

(53) Kozekov, I. D.; Nechev, L. V.; Sanchez, A.; Harris, C. M.; Lloyd, R. S.; Harris, T. M. *Chem. Res. Toxicol.* **2001**, *14*, 1482–1485.

(54) Kawanishi, M.; Matsuda, T.; Nakayama, A.; Takebe, H.; Matsui, S.; Yagi, T. *Mutat. Res. Toxicol. Environ. Mutagen.* **1998**, *417*, 65–73.

(55) Nath, R. G.; Ocando, J. E.; Chung, F.-L. *Cancer Res.* **1996**, *56*, 452–456.

(56) Sanchez, A. M.; Kozekov, I. D.; Harris, T. M.; Lloyd, R. S. *Chem. Res. Toxicol.* **2005**, *18*, 1683–1690.

(57) Niedernhofer, L. J.; Daniels, J. S.; Rouzer, C. A.; Greene, R. E.; Marnett, L. J. *J. Biol. Chem.* **2003**, *278*, 31426–31433.

(58) Amosova, O.; Coulter, R.; Fresco, J. R. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 4392–4397.

(59) Nakamura, J.; Walker, V. E.; Upton, P. B.; Chiang, S.-Y.; Kow, Y. W.; Swenberg, J. A. *Cancer Res.* **1998**, *58*, 222–225.

(60) Christner, D. F.; Frank, B. L.; Kozarich, J. W.; Stubbe, J.; Golik, J.; Doyle, T. W.; Rosenberg, I. E.; Krishnan, B. *J. Am. Chem. Soc.* **1992**, *114*, 8763–8767.

(61) Regulus, P.; Duroux, B.; Bayle, P.-A.; Favier, A.; Cadet, J.; Ravanat, J.-L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104*, 14032–14037.

(62) Chen, B.; Vu, C. C.; Byrns, M. C.; Dedon, P. C.; Peterson, L. A. *Chem. Res. Toxicol.* **2006**, *19*, 982–985.

(63) Chen, B.; Bohnert, T.; Zhou, X.; Dedon, P. C. *Chem. Res. Toxicol.* **2004**, *17*, 1406–1413.

(64) Hangeland, J. J.; De Voss, J. J.; Heath, J. A.; Townsend, C. A.; Ding, W. D.; Ashcroft, J. S.; Ellestad, G. A. *J. Am. Chem. Soc.* **1992**, *114*, 9200–9202.

(65) Angeloff, A.; Dubey, I.; Pratviel, G.; Bernadou, J.; Meunier, B. *Chem. Res. Toxicol.* **2001**, *14*, 1413–1420.

(66) Dizdaroglu, M. *Z Naturforsch B* **1976**, *31b*, 227.

(67) Kawabata, H.; Takeshita, H.; Fujiwara, T.; Sugiyama, H.; Matsuura, T.; Saito, I. *Tetrahedron Lett.* **1989**, *30*, 4263–4266.

(68) Egle, J. L.; Gochberg, B. J. *Am. Ind. Hyg. Assoc. J.* **1979**, *40*, 310–314.

(69) D'Andrea, A. D. *N. Engl. J. Med.* **2010**, *362*, 1909–1919.

(70) Guainazzi, A.; Campbell, A. J.; Angelov, T.; Simmerling, C.; Schärer, O. D. *Chemistry* **2010**, *16*, 12100–12103.

(71) Angelov, T.; Guainazzi, A.; Schärer, O. D. *Org. Lett.* **2009**, *11*, 661–664.

(72) Hentschel, S.; Alzeer, J.; Angelov, T.; Schärer, O. D.; Luedtke, N. W. *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 3466–3469.

(73) Tsarouhtsis, D.; Kuchimanchi, S.; DeCorte, B. L.; Harris, C. M.; Harris, T. M. *J. Am. Chem. Soc.* **1995**, *117*, 11013–11014.

(74) Dooley, P. A.; Tsarouhtsis, D.; Korbel, G. A.; Nechev, L. V.; Shearer, J.; Zegar, I. S.; Harris, C. M.; Stone, M. P.; Harris, T. M. *J. Am. Chem. Soc.* **2001**, *123*, 1730–1739.

(75) Ferentz, A. E.; Verdine, G. L. *J. Am. Chem. Soc.* **1991**, *113*, 4000–4002.

(76) Kawasaki, T.; Nagatsugi, F.; Ali, M. M.; Maeda, M.; Sugiyama, K.; Hori, K.; Sasaki, S. *J. Org. Chem.* **2005**, *70*, 14–23.

(77) Yoshimura, Y.; Fujimoto, K. *Org. Lett.* **2008**, *10*, 3227–3230.

(78) Fujimoto, K.; Yamada, A.; Yoshimura, Y.; Tsukaguchi, T.; Sakamoto, T. *J. Am. Chem. Soc.* **2013**, *135*, 16161–16167.

(79) Kobertz, W. R.; Essigmann, J. M. *J. Am. Chem. Soc.* **1997**, *119*, 5960–5961.

(80) Ding, H.; Greenberg, M. M. *Chem. Res. Toxicol.* **2007**, *20*, 1623–1628.

(81) Hong, I. S.; Greenberg, M. M. *J. Am. Chem. Soc.* **2005**, *127*, 10510–10511.

(82) Weng, L.; Horvat, S. M.; Schiesser, C. H.; Greenberg, M. M. *Org. Lett.* **2013**, *15*, 3618–3621.

(83) Ding, H.; Majumdar, A.; Tolman, J. R.; Greenberg, M. M. *J. Am. Chem. Soc.* **2008**, *130*, 17981–17987.

(84) Harwood, E. A.; Sigurdsson, S. T.; Edfeldt, N. B. F.; Reid, B. R.; Hopkins, P. B. *J. Am. Chem. Soc.* **1999**, *121*, 5081–5082.

(85) Wilds, C. J.; Xu, F.; Noronha, A. M. *Chem. Res. Toxicol.* **2008**, *21*, 686–695.

(86) Wilds, C. J.; Booth, J. D. M.; Noronha, A. M. *Curr. Protoc. Nucleic Acid Chem.* **2011**, *Chapter 5*, Unit5.9.

(87) Romero, R. M.; Rojsitthisak, P.; Haworth, I. S. *Arch. Biochem. Biophys.* **2001**, *386*, 143–153.

(88) Norman, D.; Live, D.; Sastry, M.; Lipman, R.; Hingerty, B. E.; Tomasz, M.; Broyde, S.; Patel, D. J. *Biochemistry* **1990**, *29*, 2861–2875.

(89) Spielmann, H. P.; Dwyer, T. J.; Hearst, J. E.; Wemmer, D. E. *Biochemistry* **1995**, *34*, 12937–12953.

(90) Drew, H. R.; Wing, R. M.; Takano, T.; Broka, C.; Tanaka, S.; Itakura, K.; Dickerson, R. E. *Proc. Natl. Acad. Sci.* **1981**, *78*, 2179–2183.

(91) Noll, D. M.; Webba da Silva, M.; Noronha, A. M.; Wilds, C. J.; Colvin, O. M.; Gamcsik, M. P.; Miller, P. S. *Biochemistry* **2005**, *44*, 6764–6775.

(92) Huang, H.; Kim, H.-Y.; Kozekov, I. D.; Cho, Y.-J.; Wang, H.; Kozekova, A.; Harris, T. M.; Rizzo, C. J.; Stone, M. P. *J. Am. Chem. Soc.* **2009**, *131*, 8416–8424.

(93) Huang, H.; Dooley, P. A.; Harris, C. M.; Harris, T. M.; Stone, M. P. *Chem. Res. Toxicol.* **2009**, *22*, 1810–1816.

(94) Dooley, P. A.; Zhang, M.; Korbel, G. A.; Nechev, L. V; Harris, C. M.; Stone, M. P.; Harris, T. M. *J. Am. Chem. Soc.* **2003**, *125*, 62–72.

(95) Leupin, W.; Chazin, W. J.; Hyberts, S.; Denny, W. A.; Wuethrich, K. *Biochemistry* **1986**, *25*, 5902–5910.

(96) Janion, C. *Acta Biochim. Pol.* **2001**, *48*, 599–610.

(97) Lamola, A. A.; Eisinger, J. *Proc. Natl. Acad. Sci. U. S. A.* **1968**, *59*, 46–51.

(98) Setlow, R. B.; Carrier, W. L. *Proc. Natl. Acad. Sci. U. S. A.* **1964**, *51*, 226–231.

(99) Seeberg, E.; Nissen-Meyer, J.; Strike, P. *Nature* **1976**, *263*, 524–526.

(100) Seeberg, E. *Proc. Natl. Acad. Sci.* **1978**, *75*, 2569–2573.

(101) Sancar, A.; Rupp, W. D. *Cell* **1983**, *33*, 249–260.

(102) Jain, V.; Hilton, B.; Patnaik, S.; Zou, Y.; Chiarelli, M. P.; Cho, B. P. *Nucleic Acids Res.* **2012**, *40*, 3939–3951.

(103) Vogel, E. W.; Barbin, A.; Nivard, M. J. M.; Stack, H. F.; Waters, M. D.; Lohman, P. H. M. *Mutat. Res. Mol. Mech. Mutagen.* **1998**, *400*, 509–540.

(104) Myles, G. M.; Sancar, A. *Biochemistry* **1991**, *30*, 3834–3840.

(105) Myles, G. M.; Hearst, J. E.; Sancar, A. *Biochemistry* **1991**, *30*, 3824–3834.

(106) Jaciuk, M.; Nowak, E.; Skowronek, K.; Tańska, A.; Nowotny, M. *Nat. Struct. Mol. Biol.* **2011**, *18*, 191–197.

(107) Pakotiprapha, D.; Inuzuka, Y.; Bowman, B. R.; Moolenaar, G. F.; Goosen, N.; Jeruzalmi, D.; Verdine, G. L. *Mol. Cell* **2008**, *29*, 122–133.

(108) Van Houten, B.; Gamper, H.; Sancar, A.; Hearst, J. *J. Biol. Chem.* **1987**, *262*, 13180–13187.

(109) Oh, E. Y.; Grossman, L. *Nucleic Acids Res.* **1986**, *14*, 8557–8571.

(110) Pakotiprapha, D.; Samuels, M.; Shen, K.; Hu, J. H.; Jeruzalmi, D. *Nat. Struct. Mol. Biol.* **2012**, *19*, 291–298.

(111) Bertrand-Burggraf, E.; Selby, C. P.; Hearst, J. E.; Sancar, A. *J. Mol. Biol.* **1991**, *219*, 27–36.

(112) Mazur, S. J.; Grossman, L. *Biochemistry* **1991**, *30*, 4432–4443.

(113) Van Houten, B.; Snowden, A. *Bioessays* **1993**, *15*, 51–59.

(114) Truglio, J. J.; Karakas, E.; Rhau, B.; Wang, H.; DellaVecchia, M. J.; Van Houten, B.; Kisker, C. *Nat. Struct. Mol. Biol.* **2006**, *13*, 360–364.

(115) Moolenaar, G. F. *J. Biol. Chem.* **2000**, *275*, 8038–8043.

(116) Orren, D. K.; Sancar, A. *Proc. Natl. Acad. Sci.* **1989**, *86*, 5237–5241.

(117) Moolenaar, G. F.; Schut, M.; Goosen, N. *DNA Repair (Amst).* **2005**, *4*, 699–713.

(118) Wang, H.; Lu, M.; Tang, M.; Van Houten, B.; Ross, J. B. A.; Weinfeld, M.; Le, X. C. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12849–12854.

(119) Malta, E.; Moolenaar, G. F.; Goosen, N. *J. Biol. Chem.* **2006**, *281*, 2184–2194.

(120) Van Houten, B.; Gamper, H.; Holbrook, S. R.; Hearst, J. E.; Sancar, A. *Proc. Natl. Acad. Sci.* **1986**, *83*, 8077–8081.

(121) Truglio, J. J.; Croteau, D. L.; Van Houten, B.; Kisker, C. *Chem. Rev.* **2006**, *106*, 233–252.

(122) DellaVecchia, M. J.; Croteau, D. L.; Skorvaga, M.; Dezhurov, S. V; Lavrik, O. I.; Van Houten, B. *J. Biol. Chem.* **2004**, *279*, 45245–45256.

(123) Moolenaar, G. F. *J. Biol. Chem.* **1998**, *273*, 34896–34903.

(124) Snowden, A.; Van Houten, B. *J. Mol. Biol.* **1991**, *220*, 19–33.

(125) McKibbin, P. L.; Fleming, A. M.; Towheed, M. A.; Van Houten, B.; Burrows, C. J.; David, S. S. *J. Am. Chem. Soc.* **2013**, *135*, 13851–13861.

(126) Zou, Y.; Liu, T.-M.; Geacintov, N. E.; Van Houten, B. *Biochemistry* **1995**, *34*, 13582–13593.

(127) Geacintov, N. E.; Cosman, M.; Hingerty, B. E.; Amin, S.; Broyde, S.; Patel, D. J. *Chem. Res. Toxicol.* **1997**, *10*, 111–146.

(128) Jones, B. K.; Yeung, A. T. *J. Biol. Chem.* **1990**, *265*, 3489–3496.

(129) Bae, S.; Lakshman, M. K. *J. Am. Chem. Soc.* **2007**, *129*, 782–789.

(130) McGall, G. H.; Barone, A. D.; Diggelmann, M.; Fodor, S. P. A.; Gentalen, E.; Ngo, N. *J. Am. Chem. Soc.* **1997**, *119*, 5081–5090.

(131) Venkatesan, H.; Greenberg, M. M. *J. Org. Chem.* **1996**, *61*, 525–529.

(132) Pirrung, M. C.; Bradley, J.-C. *J. Org. Chem.* **1995**, *60*, 6270–6276.

(133) Mascavage, L. M.; Lu, Q.; Vey, J.; Dalton, D. R.; Carroll, P. J. *J. Org. Chem.* **2001**, *66*, 3621–3626.

(134) Huwe, C. M.; Blechert, S. *Synthesis (Stuttg).* **1997**, *1997*, 61–67.

(135) Lee, J. H.; Kang, J. E.; Yang, M. S.; Kang, K. Y.; Park, K. H. *Tetrahedron* **2001**, *57*, 10071–10076.

(136) Cicchi, S.; Goti, A.; Brandi, A. *J. Org. Chem.* **1995**, *60*, 4743–4748.

(137) Merino, P.; Delso, I.; Tejero, T.; Cardona, F.; Marradi, M.; Faggi, E.; Parmeggiani, C.; Goti, A. *European J. Org. Chem.* **2008**, *2008*, 2929–2947.

(138) Hayakawa, Y.; Wakabayashi, S.; Kato, H.; Noyori, R. *J. Am. Chem. Soc.* **1990**, *112*, 1691–1696.

(139) Fernández-Forner, D.; Casals, G.; Navarro, E.; Ryder, H.; Albericio, F. *Tetrahedron Lett.* **2001**, *42*, 4471–4474.

(140) Hayakawa, Y.; Kato, H.; Uchiyama, M.; Kajino, H.; Noyori, R. *J. Org. Chem.* **1986**, *51*, 2400–2402.

(141) Ghosh, S.; Greenberg, M. M. *Biochemistry* **2014**, *53*, 5958–5965.

(142) Goutelle, S.; Maurin, M.; Rougier, F.; Barbaut, X.; Bourguignon, L.; Ducher, M.; Maire, P. *Fundam. Clin. Pharmacol.* **2008**, *22*, 633–648.

(143) Weiss, J. N. *FASEB J.* **1997**, *11*, 835–841.

(144) Maxam, A. M.; Gilbert, W. *Proc. Natl. Acad. Sci. U. S. A.* **1977**, *74*, 560–564.

(145) Bishop, E. P.; Rohs, R.; Parker, S. C. J.; West, S. M.; Liu, P.; Mann, R. S.; Honig, B.; Tullius, T. D. *ACS Chem. Biol.* **2011**, *6*, 1314–1320.

(146) Hayes, J. J.; Clark, D. J.; Wolffe, A. P. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 6829–6833.

(147) Roberts, E. *Nucleic Acids Res.* **1997**, *25*, 3377–3378.

(148) Yakubovskaya, M. G.; Belyakova, A. A.; Gasanova, V. K.; Belitsky, G. A.; Dolinnaya, N. G. *Biochimie* **2010**, *92*, 762–771.

(149) Jarem, D. A.; Huckaby, L. V; Delaney, S. *Biochemistry* **2010**, *49*, 6826–6837.

(150) Sun, D.; Hurley, L. H. *Methods Mol. Biol.* **2010**, *608*, 65–79.

(151) Rapireddy, S.; He, G.; Roy, S.; Armitage, B. A.; Ly, D. H. *J. Am. Chem. Soc.* **2007**, *129*, 15596–15600.

(152) Shcherbakova, I.; Mitra, S. *Methods Enzymol.* **2009**, *468*, 31–46.

(153) Brenowitz, M.; Senear, D. F.; Shea, M. A.; Ackers, G. K. *Methods Enzymol.* **1986**, *130*, 132–181.

(154) Greenbaum, J. A.; Pang, B.; Tullius, T. D. *Genome Res.* **2007**, *17*, 947–953.

(155) Bhattacharyya, D.; Bansal, M. *J. Biomol. Struct. Dyn.* **1990**, *8*, 539–572.

(156) Tataurov, A. V; You, Y.; Owczarzy, R. *Biophys. Chem.* **2008**, *133*, 66–70.

(157) Fasman, G. D. *Handbook of Biochemistry and Molecular Biology - Volume 1 Nucleic acids*; 3rd ed.; CRC Press, 1975.

(158) Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059–1077.

(159) Berkner, K. L.; Folk, W. R. *Polynucleotide Kinase Exchange as an Assay for Class II Restriction Endonucleases*; Methods in Enzymology; Elsevier, 1980; Vol. 65.

(160) Okazaki, R.; Hirose, S.; Okazaki, T.; Ogawa, T.; Kurosawa, Y. *Biochem. Biophys. Res. Commun.* **1975**, *62*, 1018–1024.

(161) Owczarzy, R.; Moreira, B. G.; You, Y.; Behlke, M. A.; Walder, J. A. *Biochemistry* **2008**, *47*, 5336–5353.

(162) Ghosh, S.; Greenberg, M. M. *J. Org. Chem.* **2014**, *79*, 5948–5957.

(163) Smeaton, M. B.; Hlavin, E. M.; McGregor Mason, T.; Noronha, A. M.; Wilds, C. J.; Miller, P. S. *Biochemistry* **2008**, *47*, 9920–9930.

(164) Shin, D.-S.; Lee, Y.-S. *Synlett* **2009**, *2009*, 3307–3310.

8. Curriculum Vitae

Souradyuti Ghosh was born (04/18/1985) in the town of Islampur, Uttar Dinajpur district, West Bengal state in India. Having done his high school in the same place, he completed undergraduate education with chemistry major from Ramakrishna Mission Residential College, Narendrapur (under Calcutta University) on 2006. He did his M.S. in chemistry from Indian Institute of Technology, Madras (IITM) on 2008 where his Master's thesis work was later published as a book chapter. After completing his M.S., he continued to work in IITM for another year as a project associate to study the applicability of biomacromolecules in lithium batteries. He was accepted as a doctoral student in the department of chemistry, Johns Hopkins University (JHU) at fall, 2009. In JHU chemistry, he joined Prof. Marc M. Greenberg's lab to investigate the biochemical and structural aspects of DNA interstrand cross-link (ICL) generation, a covalent linkage formed between opposing strands of DNA which hampers essential cellular processes. He designed novel strategies to synthesize ICLs and studied their repair and structure. His work in Greenberg lab has produced two first author publications with one more under preparation. He defended his thesis on March, 2015 and now works as a intern in the Biology department, JHU.

Journal Publications

1. **Ghosh, S.**; Greenberg, M. M., Synthesis of cross-linked DNA containing oxidized abasic site analogues. J. Org. Chem. 2014, 79, 5948.
2. **Ghosh, S.**; Greenberg, M. M., Nucleotide excision repair of chemically stabilized analogues of DNA interstrand cross-links produced from oxidized abasic sites. Biochemistry 2014, 53, 5958.
3. **Ghosh, S.**; Greenberg, M. M., Investigation of site selective formation of C4-AP and DOB interstrand cross-links using their stabilized analogues [in preparation for Biochemistry]

Book Chapter

1. George, C.; **Ghosh, S.**; Chandrakumar, N., ^{29}Si NMR-transition selective double quantum filter experiments in one and two dimensions. in *Future Directions in NMR*; Editors: Khetrapal**,** C. L.; Kumar**,** A.; Ramanathan**,** K. V.; Springer, 2014, pp 231-237. (ISBN 978-81-8489-995-5)